

Item analysis of the Grit Assessment Scale: International version in Portuguese language

Tatiana de Cassia Nakano*  / Ana Paula Porto Noronha**  / Ligia de Santis**  / Leandro S. Almeida*** 

* Pontifícia Universidade Católica de Campinas, Campinas, Brasil; ** Universidade São Francisco, Campinas, Brasil; *** Universidade do Minho, Braga, Portugal

Abstract: Grit can be understood as an element of personality, being a positive and non-cognitive human characteristic, composed of two components: perseverance of effort and consistency of interests. Due to its social, professional, and educational impact, grit has become a construct of great interest. Although some instruments have been developed and adapted for its evaluation in different contexts, the absence of measures in Portuguese led to the development of the Grit Assessment Scale – International Version in Portuguese Language (EAGrit-LP), based on samples from Brazil and Portugal. Even though previous studies have investigated some evidence of validity for the instrument, its items had not yet been individually evaluated. Therefore, the objective of the present study was to perform an item analysis, using Item Response Theory. For this purpose, a sample of 1.050 Brazilians and 656 Portuguese individuals responded to the EAGrit-LP (aged between 17 and 71 years old; $M = 24.1$ years; $SD = 9.2$). The results showed that all items had appropriate infit and outfit indices, and that items difficulty was, in general, low. Six items displayed differential item functioning (DIF), based on participants' gender and country of origin. These results can be used to revise the scale and guide the development of more challenging items, as well as to the development of normative tables that are specific to gender and country of origin.

Keywords: Positive psychology, Psychometrics, Tests construction, Psychological assessment.

Grit originates from positive psychology (Schmidt et al., 2021) and was initially proposed by Duckworth et al. (2007), as a fusion of passion and perseverance in the pursuit of long-term goals. This construct consists of two components: perseverance of effort (PE) and consistency of interests (CI) (Duckworth & Quinn, 2009). Essentially, PE refers to maintaining effort and reaching long-term goals as well as tendencies to continue working (Duckworth et al., 2021). When relating to CI, it refers to the tendency to maintain focus over a long period of time when achieving goals, successfully capturing the notion of passion. It is important to know that CI reflects a disposition concerning a particular topic that is not limited to a particular situation (Muenks et al., 2017). Based upon the authors' model of two factors, PE and CI are interrelated, but distinct (Kuruveettissery et al., 2023).

Over the past years, grit has gained exponential interest across various fields (e.g., politics, healthcare, research and media) (Muenks et al., 2017) and a rigorous measurement of grit has been encouraged in fields like academics, careers, and lifestyle (Duckworth & Quinn, 2009). Recent studies indicated that grit plays an important role in predicting success, years of schooling, work performance, failure, health outcomes, sports success, and school performance (Park et al., 2020; Pendyala & Vyas,

2023). Furthermore, grit has been studied in the context of military service, relationships, workplace, and professional skills (Reysen et al., 2019), and it seems to be predictor of happiness in study situations or subjective well-being (Zhang et al., 2023), particularly through the PE dimension. Grit has been shown to be a malleable and promising personality trait, it can also be developed through deliberate practice (Schmischal et al., 2022), which suggests that it may be appropriate to build efforts towards the development and improvement of instruments to support interventions.

This model has been widely studied in different educational settings (Duckworth et al., 2021; Jiang et al., 2021; Morell et al., 2021; Muenks et al., 2017). Investigations have found associations between PE and the use of learning strategies (Jiang et al., 2021), self-efficacy levels (Muenks et al., 2017), grades (Morell et al., 2021), and years of education completed by adults (Duckworth et al., 2021). High levels of CI are associated with positive effects on student organization (Jiang et al., 2021). Moreover, grit is a relevant predictor of academic and vocational outcomes, such as fewer career changes and greater educational attainment (Schmidt et al., 2021).

In educational context, CI explains why some individuals spend a substantial amount of time, effort, and hard work to reach an achievement, even if the results are not immediate (Neroni et al., 2022). The concept of PE may be understood as a commitment, involvement and effort that drives us towards achieving something (Sigmundsson et al., 2020). Due to this specificity of meaning of the two dimensions, PE, but not CI, is shown to be a predictor of students' academic results (Neroni et al., 2022). Moreover, grit is considered a performance character strength by scholars, showing moderate correlations with educational variables like academic performance and positive variables such as hope, positive affect, and family relationships (Hasan et al., 2022).

The environment plays an important role in developing grit, just as personality traits do. In this regard, Teimouri et al. (2021) assert that researchers have sought to measure this capability across a wide range of areas, especially in the fields of sport and education. Given that grit is considered a specific domain, individuals can demonstrate different levels of passion (or consistency of interest) and perseverance (Cormier et al., 2019). Thus, a person may display grit in his professional life, but not in his personal life. Furthermore, the authors point out that, although there may be some stability between domains, certain variables, such as cultural context, social role, gender, and language tend to differentiate levels of competence (Cormier et al., 2019).

As a result, it cannot be denied that culture plays an important role in influencing individual behavior, especially in its relation to passion, and more generally in its relation to perseverance (Kuruveettissery et al., 2023). Its manifestation may be affected by cultural nuances of the individuals in a particular cultural context (Hasan et al., 2022). It is crucial to understand the cultural interpretations about what means to have a passion, sustain it for long periods, and persevere to succeed (Pendyala & Vyas, 2023).

Given its relevance, some grit assessment instruments have been developed. In view of this, Hasan et al. (2022) revised the literature on grit scales that presents psychometric studies. After the literature review, they pointed the existence of several grit scales: the Original Grit Scale (Grit-O), the Short Grit Scale (Grit-S), the Triarchic Model of the Grit Scale (TMGS), the Grit Psychological Resources Scale (GPRS) (Schmischal et al., 2022), the Grit Scale for Children and Adults (GSCA), the Academic Grit Scale (AGS), the L2-Teacher Grit Scale (Sudina et al., 2020), the Oviedo Grit Scale (EGO) (Postigo et al., 2021), the Clinical Nurses Grit Scale (CN-Grit) (Park et al., 2020), the 12-item Multidimensional Grit Scale (Pajestka & Poraj-Weder, 2023), 3-D Grit Scale (Kuruveettissery et al., 2023), Physical Education Grit Scale (PE-Grit Scale) (Guelmami et al., 2022), Domain-specific Grit Scale for College Athletic Students (DGSCAS) (Gao et al., 2024), English as a Foreign Language Grit Scale (EFL) (Ebadi et al., 2018), Sports Grit Scale (Fruchart & Rulence-Pâques, 2024), Grit Scales and the Long-term Grit Scales (L2 Grit Scale) (Li & Yang, 2023), and the Metacognitive Awareness of Grit Scale (MCAGS).

Although there are many instruments on grit, in recent years, two scales have gained significant popularity, the Grit Scale (Grit-O) (Duckworth et al., 2007) and the Grit Scale Short (Grit-S) (Duckworth et al., 2021). The first one included 12 items, six related to consistency of interests (such as “New ideas and projects sometimes distract me from previous ones”, a negative item) and six to perseverance of effort (such as “I am a hard worker”). The short version of the scale (the Grit-S) has eight-item (four fewer items than the Grit-O) and improved psychometric properties (Duckworth et al., 2021).

Considering that grit manifestation may not be universal, varying among settings and cultures (Hasan et al., 2022), it is necessary to create or validate grit scales across diverse cultures (Li et al., 2018). Faced with that, many studies have translated the Grit-O into different languages and contexts (e.g., Areepattamannil & Khine, 2018; Schmidt et al., 2021; Tyumeneva et al., 2014; and others), including cross-cultural studies (e.g., van Zyl et al., 2022). However, there are no versions of this scale that can be applied to the Brazilian context.

Considering instruments in Portuguese (suitable for assessing Brazilians and other speakers of the language, such as Portugal) there is only one measure for assess grit in Portuguese, the Oviedo Grit Scale. However, limited research has been conducted on its psychometric properties (only internal consistency, factor structure, and invariance considering gender and curricular year) (Monteiro et al., 2023). Also, the fact that the Oviedo Grit Scale was only tested with people from Portugal (not with Brazilians) limits its use in Brazil.

Considering this lack of measures to evaluate grit in the Brazilian context, Noronha and Almeida (2022) developed the Grit Assessment Scale – International Version in Portuguese Language (EAGrit-LP), based on Brazilian and Portuguese data. Given this sample composition, the measure enables grit evaluation in Brazil and Portugal, making it possible to conduct intercultural evaluations on the topic. In addition to these advantages, the measure was developed considering international results on other grit instruments, which indicated inconsistencies in its internal structure (as can be seen in the results of these studies: Arco-Tirado et al., 2018; Areepattamannil & Khine, 2017; Beri & Sharma, 2019; Marentes-Castillo et al., 2019; Schmidt et al., 2021; Sordia, 2020; Tan et al., 2019).

To demonstrate their psychometric qualities, psychological instruments used in research and intervention must undergo a rigorous construction process. Among these processes are the estimation of reliability and the investigation of validity evidence (American Educational Research Association [AERA] et al., 2014). Identifying whether the test measures a psychological construct and whether it replicates the theoretical framework that underlies its development is one of the most challenging and critical tasks (Nakano et al., 2015). In this sense, during the development of an instrument, item analysis plays a crucial role since it allows each item to be evaluated separately, contributing significantly to the interpretation of its scores (Bock & Gibbons, 2021; Wright & Masters, 1982). To verify evidence of this nature, estimates derived from Item Response Theory (IRT) are relevant. IRT can provide more informative parameters (like item-difficulty, person ability, and item-discrimination parameters) that are not sample-dependent, reflecting accurate estimates of misfit items (Alhadabi, 2023). Also, the differential item functioning can help to understand the variance among different groups (for example, two different cultures).

In this context, several studies have adopted IRT-based approaches to investigate the psychometric properties of Grit scales. Analyses of this nature were carried out, for example, for the Grit-S Scale (Alhadabi, 2023; Areepattamannil & Khine, 2017; Farmer et al., 2025; Gonzalez et al., 2020; Midkiff et al., 2017; Morell et al., 2021; Muenks et al., 2017; Tyumeneva et al., 2019; Yupanqui-Lorenzo et al., 2024), the NL-Grit scale (Dijksma et al., 2023), and the Grit scale – GS-12 (Jie et al., 2024). A reason for this interest can be explained by Credé et al. (2017) who proposed that grit literature may benefit significantly from the refinement of its scales based on item response theory. In addition, IRT provide useful information about the contribution of the items in the

measurement of the latent construct, its quality, and at which points the ability scale performs the best measurement, after places examinees and items on the same scale (Yiğiter & Boduroğlu, 2024).

Farmer et al. (2025) assert that the IRT provides further information as it calculates the accuracy of a scale within a range of a latent trait, “describe the relationship between an individual’s level on the construct (i.e., theta) and the probability of a particular response occurring, as well as describing the difficulty of each item, and it can incorporate this information in scaling” (p. 2). It has been suggested by Jie et al. (2024) that IRT provides a more accurate assessment of an individual’s ability level, allowing a more precise description of measurement characteristics.

Considering that the EAGrit-LP is the only grit measure developed for Brazilians, it is important to understand its psychometric qualities. Preliminary studies have already verified validity evidence based on its content (Noronha & Almeida, 2022) as well as based on its internal structure (Noronha et al., 2024). Results indicated that the measure presents a multidimensional internal structure, composed of two factors (Pe and CI), with excellent reliability (Noronha et al., 2024). However, the quality of its items, individually, have not yet been evaluated. Thus, the aim of the present study was to analyze the EAGrit-LP items to gain a deeper insight into the relationship between the ability (latent traits) and the test results (Bock & Gibbons, 2021). For that purpose, a Rasch model (Bond & Fox, 2015) was used to assess fit indices (infit and outfit) and item difficulty, as well differential item functioning based on participants’ gender and country of origin. There have been many studies conducted on the use of DIF in Grit scale analysis, aiming to compare two cultures, including an analysis of grit scale by Alhadabi (2023), across racial groups by Gladstone et al. (2022), across genders and grade levels by Areepattamannil and Khine (2017), and gender-based by Lin and Shih (2021).

Method

Participants

A total of 1.706 participants from a convenience sample were included in the study: 1.050 Brazilians and 656 Portuguese, aged between 17 and 71 years old ($M = 24.1$ years; $SD = 9.2$). The sample consisted mainly of females (72.1%) and of students attending private universities (51.0%). Also, most students were in the initial semesters of their studies (22.0%), followed by 48.5% in the middle semesters, and 29.5% in the final semesters.

Instrument

Sociodemographic questionnaire. The questionnaire was developed for the present study. It aimed to access sociodemographic information including country of residence, gender and age.

Grit Assessment Scale – EAGrit-LP (Noronha & Almeida, 2022; Noronha et al., 2024). A 12-item scale that evaluates two factors: consistency in interests and perseverance of effort, each with six items. The instrument is answered on a Likert scale with four possible responses: 1 = completely disagree, 2 = sometimes disagree, 3 = sometimes agree, and 4 = completely agree. Based on the responses to the items that make up each factor, scores may vary between six and twenty-four, with higher scores indicating higher levels of grit. Developed based on Brazilian and Portuguese data, the EAGrit-LP has been the subject of studies focusing on its construction and evidence of content validity (Noronha & Almeida, 2022), as well as a study investigating validity evidence based on the factorial structure, with adequate results (Noronha et al., 2024).

Procedures

The cross-sectional study was performed between May and November of 2022 via google forms, including the Free and Informed Consent Form on the first page, the Sociodemographic Questionnaire and the EA Grit-LP. In Brazil, the questionnaire was disseminated through social networks (Instagram, Facebook, and LinkedIn), as well as via email. In Portugal, the questionnaire was shared in classrooms at Higher Education institutions.

Data analysis

Initially, descriptive statistics were calculated. To comply with the most popular assumption of IRT models, the unidimensionality, as a way to guarantee that all items of the test measure the same latent trait (Gyamfi & Acquaye, 2023) and only one characteristic of the subject (Hu et al., 2021), item analysis was conducted separately for each of the factors – perseverance of effort (PE) and consistency of interests (CI). To address the polytomous nature of the items, the Rasch-Andrich Rating Scale Model (Andrich, 1978) has been used on JAMOVI software 2.6.44. The result tables are estimated by Marginal Maximum likelihood estimation (MMLE) and the eRm, a R Package was used for the person-item map.

For the purposes of evaluating the items, item difficulty (parameter b) and fit indices (infit and outfit) were estimated. The item difficulty index is related to the probability of endorsing an item and reflects the level of ability required to agree with the item's content (Gyamfi & Acquaye, 2023). To calculate the average amount of ability (θ) required for the item to be endorsed, the reliability of individuals was assessed, and maps of the item/construct were constructed. The Item-People Map procedure was employed to determine the relationship between the difficulty of the items and the ability levels presented by the participants (Embretson & Reise, 2000). This value usually ranges between -2.5 and +2.5, in theory (Ahangari & Semiyari, 2022).

Items suitability for the model were also evaluated. This analysis was carried out by estimating fit indices, a procedure that allows detecting differences between what was predicted by the model and what was empirically observed in the data (Smith, 2004; Wright & Masters, 1982). These differences are called residuals, and their evaluation includes two main indices: infit (to check for discrepancies in items whose difficulties are close to person's abilities level) and outfit (to check for discrepancies in extreme items).

Assessing item fit is a central point in IRT, considering that a good fit is necessary to draw valid inferences from estimated model parameters (Diaz et al., 2022). This is true because the fit indexes can identify examinees that exhibit patterns of unexpected responses. Infit and outfit values can range between zero and infinity (Jones et al., 2023). According to Linacre and Wright (2002), however, values below and above 0.5 and 1.5 (especially for infit) indicate a mismatch between the model and observed data.

Subsequently, we estimated the differential item functioning (DIF), based on the gender and country of origin of the participants. This type of analysis is used to verify the existence of items with different difficulty between groups, favoring one of them. In this sense, depending on the group where the person is inserted, the item may be more easily endorsed or require a higher level of skill to be endorsed (Yiğiter & Boduroglu, 2024). Items that present DIF indicate that individuals who have the same level of ability can perform differently on a test, revealing that these items cannot assess the construct fairly, because the assessment would depend on the group to which the person belongs (Ahangari & Semiyari, 2022). In the present study, as a criterion for assessing DIF, values of probability below 0.05 or DIF contrast greater than 0.42 was adopted, following literature guidelines (Aguerri et al., 2007).

Results

Initially, the results are presented for Factor 1 (Perseverance of Effort), consisting of six items. Descriptive statistics indicated that participants' scores varied between 7 and 24 points ($M = 19.0$; $SD = 3.42$). The person reliability was 0.80. Regarding the difficulty of each item, as well as infit and outfit, the results are presented in Table 1.

Table 1

Factor 1 – Perseverance of effort: Item difficulty and fit index

	<i>b</i>	<i>SE</i>	<i>SD</i>	Infit (MNSQ)	Outfit (MNSQ)
Item 2 – I persevere in my efforts to achieve my goals	-4.34	0.05	0.04	1.04	1.07
Item 6 – Even if some important activity is difficult, I insist until I manage to do it	-4.19	0.05	0.04	0.97	0.94
Item 8 – I follow my plans, even if I have adversity to execute them	-3.94	0.01	0.04	0.95	0.95
Item 10 – If I start a task, I focus on its completion	-3.99	0.05	0.04	1.04	1.00
Item 11 – I persist to achieve my goals	-4.44	0.05	0.04	0.90	0.82
Item 12 – When I start a project, I can see it through to the end.	-4.13	0.05	0.04	1.07	1.03

Note. *SD* = standard deviation; *b* = item difficulty; *SE* = standard error; *MNSQ* = mean square error.

The data indicate that the difficulty of the items was quite low, so that a high level of ability (grit) is not necessary for the person to endorse the highest response alternatives. None of the items in this factor showed a mismatch in the residuals, as values for infit and outfit are adequate.

In accordance with the construct map (Figure 1 on the left), most participants have an ability level of -1, ranging from -6.5 to +3.0. Regarding the items, the difficulty analysis indicates that item 2 and item 11 are the easiest to endorse, while item 8 and item 10 require more ability to endorse, despite the fact that the difficulty of all items are quite low. When ability level and response options are analyzed for all items, individuals with theta of one already choose alternative 2 of the response scale (Figure 1, on the right). As can be seen in the curve showing the distribution of participants at the top of the figure, participants have a greater tendency to mark alternative 3.

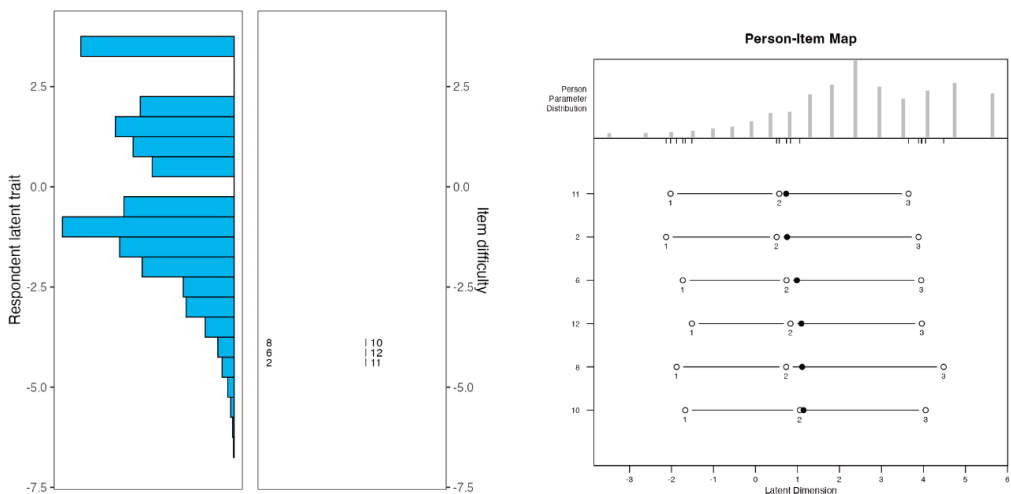


Figure 1. Construct map for Factor 1 (perseverance of effort)

To identify differences due to participants' gender and country of origin, differential item functioning (DIF) was estimated. Table 2 presents the results for Factor 1 (Perseverance of Effort).

Table 2

DIF for country of origin and gender for Factor 1 – Perseverance of effort

DIF for country of origin	Brazil	Portugal	Dif contrast	Probability
Item 2	-0.09	-0.26	0.17	0.10
Item 6	-0.07	-0.02	-0.05	0.51
Item 8	-0.02	0.39	-0.42	0.01*
Item 10	0.17	0.22	-0.04	0.52
Item 11	-0.18	-0.39	0.21	0.34
Item 12	0.19	0.00	0.19	0.10
DIF for gender	Female	Male	Dif contrast	Probability
Item 2	-0.09	-0.26	0.7	0.10
Item 6	-0.07	-0.02	-0.05	0.51
Item 8	-0.02	0.39	-0.42	0.01*
Item 10	0.17	0.22	-0.04	0.52
Item 11	-0.18	-0.39	0.21	0.34
Item 12	0.19	0.00	0.19	0.10

Note. DIF = different item functioning.

The results indicated that, in relation to the country of origin, only one item (item 8: “I follow my plans, even if I have adversity to execute them”) presented DIF, requiring a lower level of ability to be endorsed by Brazilian students (compared to Portuguese students). In relation to gender, the same item (item 8) presented DIF, favoring female participants (compared to male).

The same analyzes were then conducted with the six items that made up Factor 2 (Consistency of Interests). Results are presented in Table 3, regarding items difficulty and adjustment indices. The person reliability for this factor was 0.80.

Table 3

Factor 2 – Consistency of interests: Item difficulty and fit index

	<i>b</i>	<i>SE</i>	<i>SD</i>	Infit (MNSQ)	Outfit (MNSQ)
Item 1 – I have my life goals set for the next few years	-3.30	0.04	0.04	0.91	0.91
Item 3 – I try to commit to achieving my goals	-4.76	0.04	0.04	1.12	1.07
Item 4 – My interests are stable	-3.43	0.04	0.04	1.10	1.09
Item 5 – I have defined what I intend to do in my life	-3.43	0.04	0.04	0.97	0.95
Item 7 – I maintain a coherent line of goals over time	-3.43	0.04	0.04	0.91	0.91
Item 9 – Whenever I think about my goals, I feel that they are increasingly consolidated	-3.08	0.04	0.04	1.00	1.01

Note. *SD* = standard deviation; *b* = item difficult; *SE* = standard error; *MNSQ* = mean square error.

As with Factor 1 (Perseverance of Effort), the results for CI factor indicate that the difficulty of the items is not particularly high, showing that little ability (grit) is required by the person to endorse the content of the items. Regarding fit indices, both infit and outfit results were adequate (Table 3).

The results of participants' ability level and item difficulty are presented in Figure 2 (left figure). It is possible to verify that most participants have ability level close to theta of -1.0. In terms of the items, the item 3 is the easiest to endorse (less ability required), while item 9 requires the highest level of ability for its endorsement.

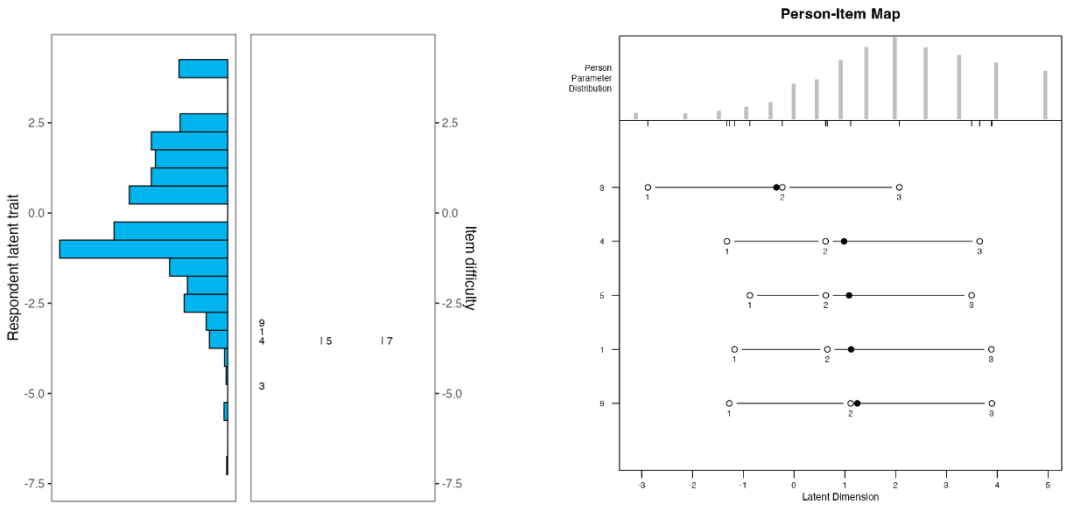


Figure 2. Construct map for Factor 2 (consistency of interests)

As in Factor 1, the relationship between ability level and response options indicates that, for all items, a person with theta of one already selects alternative 2 of the response scale (as can be seen in Figure 1 on the right). According to the curve that illustrates the distribution of people at the top of the figure, participants tended to choose alternative 3 more frequently.

Additionally, the differential item functioning (DIF) was examined for the second factor, considering the variables gender and country of origin. Results are presented in Table 4.

Table 4

DIF for country of origin and gender for Factor 2 – Consistency of interests

DIF for country of origin	Brazil	Portugal	Dif contrast	Probability
Item 1	0.33	0.33	0.00	0.87
Item 3	-1.85	-1.21	0.64	0.00*
Item 4	0.21	0.15	-0.07	0.95
Item 5	0.24	0.12	-0.12	0.03*
Item 7	0.22	0.14	-0.07	0.41
Item 9	0.71	0.50	-0.21	0.08
DIF for gender	Female	Male	Dif contrast	Probability
Item 1	0.17	0.39	-0.22	0.02*
Item 3	-1.13	-1.55	0.42	0.00*
Item 4	0.00	0.24	-0.25	0.03*
Item 5	0.17	0.17	0.00	0.78
Item 7	0.32	0.11	0.22	0.02*
Item 9	0.50	0.61	-0.11	0.29

Note. DIF=differential item functioning.

Among the items relating to the CI factor, there were differences in performance by country of origin for item 3 (“I try to commit to achieving my goals”), favoring Brazilian respondents, and for item 5 (“I have defined what I intend to do in my life”), favoring Portuguese respondents. These results indicate that both items have different probabilities of endorsement among individuals who have the same level of ability but belong to a different group (in this case, country of origin).

Regarding gender, four out of six items presented DIF. Item 1 (“I have my life goals set for the next few years”) and item 4 (“My interests are stable”) showed to be easier to endorse by female participants (compared to male). By their turn, item 3 (“I try to commit achieving my goals”) and item 7 (“I maintain a coherent line of goals over time”) required a lower level of ability to be endorsed by male participants (compared to female).

Discussion

The present study intended to investigate the psychometric properties of the EAGrit-LP items (Noronha & Almeida, 2022; Noronha et al., 2024), particularly assessing the relationship between latent trait levels of the individuals and test results (Bock & Gibbons, 2021), by analyzing fit indices (infit and outfit), item difficulty, and differential item functioning based on participants’ gender and country of origin. IRT-based estimates can be useful for identifying areas with potential for improvement in psychometric instruments, such as poorly performing items that cannot distinguish between response levels, items with low reliability, and items that present increased redundancy (Farmer et al., 2025).

The EAGrit-LP was developed based on Duckworth et al. (2007)’s grit comprehension. The results of an exploratory and confirmatory factor analysis indicated that the scale had a two-factor structure and that its factor structure was invariant between country of origin (Portugal and Brazil), as well as between respondents’ gender (male and female) (Noronha et al., 2024). Although additional analyses were still needed (such as the assessment of the relationship between latent trait of individuals and their responses to the instrument), these results already provided a valuable contribution regarding the psychometric properties of the instrument (Nakano et al., 2015) – indicating the potential of the EAGrit-LP to assess grit in Portuguese language. A rigorous measure of grit is essential for quantifying perseverance of effort and consistency of interest in various areas, as Kuruveettissery et al. (2023) highlight.

In the present study, by using Item Response Theory, we have been able to better understand the functioning of the scale and its items. Since most psychometric studies of grit scales investigate validity (especially based on the internal structural of the measure and on the relationship with external variables), reliability estimates, and the association with success outcomes, this study is distinguished by providing different evidence. For example, when items are analyzed on other scales, only the item total correlation is considered (Hasan et al., 2022). Thus, Tynan (2021) recommended the use of Item Response Theory for the development and evaluation of measures. As determined by the analysis of residuals (infit and outfit), all items that constitute the EAGrit-LP presented adequate results. Thus, the model can predict both the results of individuals with ability levels close to the difficulty of the item and those with extreme values.

Regarding the items difficulty (i.e., the level of latent trait required for a given response to be given on the scale), results on the EAGrit-LP showed that, in general, the values on the latent trait required to endorse the items are low. These results imply that the items present content (or situations) that can be readily endorsed by individuals who possess low levels of ability (for the present study, grit). This result was shared by both factors of the instrument. Regarding item

difficulty, item 11 (“I persist to achieve my goals”; $b = -4.44$) in PE and item 3 (“I try to commit to achieving my goals”; $b = -4.76$) in CI represented the easiest items. Considering (a) the difficulty results of all items (ranging from -3.99 to -4.44 for the PE and from 3.08 to -4.76 for the CI) and (b) that the average ability level of individuals in the present sample was $\theta = -1$, we can verify that this average ability level probably causes all items to be endorsed by the participants, selecting the higher response alternatives on the EAGrit-LP’s four points Likert scale (3 – “I” sometimes agree” or 4 – “I totally agree”).

Gonzalez et al. (2020) reported similar findings regarding the Grit-S, showing that the instrument mostly distinguished participants that were at or below the mean trait on the grit construct. As a rule, a range of -2,00 to +2,00 is considered optimal for item difficulty (Alhadabi, 2023). In the current study, the values are lower than this criteria, and one item displayed underfitting to the Rasch model based on Areepattamannil and Khine (2017). The results of Areepattamannil and Khine (2017) study using the Original Grit Scale were similar with all items showing good fit indices, except for one item.

Furthermore, we determined the differential item functioning (DIF) based on the gender and country of origin of the participants. Based on the DIF results, it is possible to identify items that behave differently in different groups. Results revealed that three items presented DIF for country of origin and five items for gender, these items being more frequent for the Consistency of Interests factor. It is understood that the probability of endorsing a given item presenting DIF is different between the groups analyzed (Fidalgo & Scalón, 2012) – in the present case, gender and country of origin. These results indicate that other variables are influencing the results of the test, beyond the respondent’s grit levels.

In studies using IRT in other grit scales, mixed results have been reported. In Tyumeneva et al. (2014) study, the items of Grit Scale did not demonstrate significant DIF when comparing different genders. No significant DIF was also found in Areepattamannil and Khine (2017) study, when comparing participants’ grade level and gender. In contrast, DIF was found in some studies. Alhadabi (2023), for example, found DIF in two items of the Grit Scale, when comparing Omani and American samples. Midkiff et al. (2017) also found DIF in one item when comparing first-generation college students with non-first-generation college students. Finally, Matore et al. (2023) found DIF in one item in their comparison between gender, favoring males over females.

When significant DIF results are found, it can be derived that results on the item can favor or harm the performance of one group over another. This happens because, for these items, individuals with the same level of ability in the construct have different probabilities of endorsing them simply because they belong to groups with distinct characteristics (Nakano et al., 2015). That is, in these cases, differences in results for the item are due to characteristics other than the level of the latent trait being assessed. Thus, according to Nakano et al. (2015), studies that assess DIF can identify influences that a second dimension, related to characteristics of subgroups, exerts on the instrument’s items, altering their difficulty.

Regarding results for the EAGrit-LP, it is noted the significant number of items with DIF related to gender. These results indicate the importance of taking measures so that test results can be interpreted fairly for both groups (male and female). Considering that DIF is an analysis that aims to ensure that items do not favor one subgroup when compared to another (Ackerman & Ma, 2024), the results can help to ensure fairness and equity for all examinees (Jafaripour et al., 2024). Because of this, it is important that DIF is verified during the test construction process, helping researchers to support more accurate decisions related to items (Ahangari & Semiyari, 2022), considering possible disparities caused by variables unrelated to the construct being assessed.

In the case of the present study, when scoring and interpreting items that present DIF, there is a chance that the interpretation of grit based on the results of different groups (men or women and Brazilian and Portuguese individuals) is misleading, as certain items do not have the same

probability of reaction (Fitriani & Situmorang, 2023). To avoid misinterpretation, the authors suggested that the instrument should be revised first, particularly in terms of content to further improve this reliability and trustworthiness.

In this sense, different procedures are proposed in the literature to minimize the influence of items that present DIF on the instrument score – even if there is no consensus on the most appropriate one (Sass, 2011). One possible action is the elimination of items that are subject to bias (Byrne et al., 1989). This option is indicated mainly when there are sufficient items to assess the construct of interest, covering an adequate range of variation of the latent trait (Peixoto et al., 2019). Considering the results of the present study, it is noted that this exclusion could harm the assessment of the latent trait (even more so when considering that of the items with DIF for gender, four of them are from the CI factor, and the exclusion would be unfeasible considering the number of items in the factor).

Therefore, it is suggested that the source of this bias be overcome before establishing normative tables for the interpretation of EAGrit-LP scores. It appears that the results of investigations of gender differences in the grit assessment are contradictory, sometimes suggesting differences in favor of one gender over the other, and sometimes suggesting no differences at all (Postigo et al., 2021). There may be gender differences in the dimensions of grit; however, these differences cannot be influenced by differences in responses to the test items. With this normative action, possible differences would not result in an assessment that would disfavor some of the evaluated groups.

The importance of grit is attributed to its potential for use in human resource management, recruitment, development, and training, as well as its implications for educators and practitioners when designing interventions for noncognitive development. Therefore, it is imperative to develop measures to adequately evaluate grit. In this sense, the results of the present study, on the psychometric properties of the EAGrit-LP 's items, added to the previous results on its evidence of validity, attesting the potential use of the instrument. In this context, this Portuguese language grit scale (EAGrit-LP) can contribute to reducing the scarcity of measures of the construct for Brazil and Portugal.

Final considerations

This study provides additional evidence regarding the psychometric characteristics of the Grit Assessment Scale – International version in Portuguese (EAGrit-LP). Based on the analysis of the instrument items, results of the present study revealed the relevance of the measure, as well as some characteristics that could be modified or improved. For example, depending on the interest in using the instrument, items with a greater level of difficulty could be included to adequately assess the different levels of grit in the adult populations of both countries (Brazil and Portugal). Furthermore, the existence of differential item functioning by gender and country of origin indicates the importance of developing separate normative tables for each of the groups, when preparing the standardization of the scale. That would be indicated to avoid possible biases in the correction and interpretation of data derived from the EAGrit-LP use.

Despite all these conclusions, this investigation entails certain limitations that must be addressed. Due to the differences in sample sizes between the two countries, with a sample primarily composed of Brazilians, caution should be exercised when interpreting the present results, particularly when considering the findings regarding DIF based on country of origin. For the instrument to be useful in psychological assessment in both countries, future studies should investigate other psychometric properties of the measure, considering a more representative sample

of both countries. Also, it would also be interesting to investigate other psychometric characteristics of the EAGrit-LP, especially those usually investigated for other grit scales, like its relationship with success outcomes. In addition, future studies may investigate grit in clinical populations given that most studies have examined student populations (Datu, 2020).

Funding

This work is funded by Fundação de Amparo à Pesquisa do Estado de São Paulo, Brasil (2022/06062-7).

Informed consent

All participants in this study were informed of the purpose of the study and how data will be used. They were assured that their identities would remain anonymous across the study.

Declaration of conflicting of interests

The author(s) declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval

The study was approved by the Ethical Committee of Universidade São Francisco with the reference number 5.264.572.

Author contributions

Conceptualization: APPN, LSA; Data curation: APPN, LSA; Formal analysis: TCN; Funding acquisition: APPN; Methodology: APPN, LSA; Writing – Original draft: APPN, LS; Writing – Review and edit: LS.

All authors read and approved the final manuscript.

References

- Ackerman, T. A., & Ma, Y. (2024). Examining differential item functioning from a multidimensional IRT perspective. *Psychometrika*, *89*, 4-41. <https://doi.org/10.1007/s11336-024-09965-6>
- Aguerri, M. E., Galibert, M. S., Lozzia, G. S., Abal, F. J. P., & Attorresi, H. F. (2007). Medida del funcionamiento diferencial del ítem en el marco de la teoría de respuesta al ítem. *Interdisciplinaria*, *24*(1), 95-110. <https://www.scielo.org.ar/pdf/interd/v24n1/v24n1a04.pdf>
- Ahangari, S., & Semiyari, S. R. (2022). Examining differential item functioning (DIF) for Iranian EFL test takers with different fields of study. *Research in English Pedagogy*, *10*(1), 169-190. <https://doi.org/10.30486/RELP.2021.1935588.1295>
- Alhadabi, A. (2023). Psychometric properties of Grit-S Scale in university students: An application of the graded response model. *Cogent Education*, *10*(1), 2162699. <https://doi.org/10.1080/2331186X.2022.2162699>

- American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561-573. <https://doi.org/10.1007/BF02293814>
- Arco-Tirado, J. L., Fernández-Martín, F. D., & Hoyle, R. H. (2018). Development and validation of a Spanish version of the Grit-S Scale. *Frontiers in Psychology*, *9*(96). <https://doi.org/10.3389/fpsyg.2018.00096>
- Areepattamannil, S., & Khine, M. S. (2017). 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. <https://doi.org/10.1177/0734282917719976>
- Beri, N., & Sharma, A. (2019). An evaluative study of reliability and validity of Grit 12 item scale in Indian context. *Journal of Indian Association for Child & Adolescent Mental Health*, *15*(3), 48-60. <https://doi.org/10.1177/0973134220190304>
- Bock, R. D., & Gibbons, R. D. (2021). *Item response theory*. Wiley & Sons.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge/Taylor & Francis Group.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456-466. <https://doi.org/10.1037/0033-2909.105.3.456>
- Cormier, D. L., Dunn, J. G. H., & Dunn, J. C. (2019). Examining the domain specificity of grit. *Personality and Individual Differences*, *139*, 349-354. <https://doi.org/10.1016/j.paid.2018.11.026>
- Credé, M., Tynan, M. C., & Harms, P. D. (2017). Much ado about grit: A meta-analytic synthesis of the grit literature. *Journal of Personality and Social Psychology*, *113*(3), 492-511. <https://doi.org/10.1037/pspp0000102>
- Datu, J. A. D. (2020). Beyond passion and perseverance: Review and future research initiatives on the science of grit. *Frontiers in Psychology*, *11*, 545526. <https://doi.org/10.3389/fpsyg.2020.545526>
- Díaz, J. A. S., Köhler, C., & Hartig, J. (2022). Performance of infit and outfit confidence intervals calculated via parametric bootstrapping. *Applied Measurement in Education*, *35*(2), 116-132. <https://doi.org/10.1080/08957347.2022.2067540>
- Dijkma, I., Stuiver, M., Lucas, C., & Lindeboom, R. (2023). Evaluating the psychometric properties of the Grit Scale in Marine recruits using Rasch analysis. *BMJ Military Health*, *169*, 425-429. <https://militaryhealth.bmj.com/content/jramc/169/5/425.full.pdf>
- 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. <https://doi.org/10.1037/0022-3514.92.6.1087>
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment*, *91*(2), 166-174. <https://doi.org/10.1080/00223890802634290>
- Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2021). Revisiting the factor structure of grit: A commentary on Duckworth and Quinn (2009). *Journal of Personality Assessment*, *103*(5), 573-575. <https://doi.org/10.1080/00223891.2021.194022>
- Ebadi, S., Weisi, H., & Khaksar, Z. (2018). Developing an Iranian ELT context-specific grit instrument. *Journal of Psycholinguistic Research*, *47*(4), 975-997. <https://doi.org/10.1007/s10936-018-9571-x>
- Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

- Farmer, A. Y., Wei, Y., Gale, A., & Peterson, N. A. (2025). IRT analysis of the Grit-S Scale: Evaluation with racially/ethnically diverse adolescents in the US. *Journal of Psychoeducational Assessment*, 43(2). <https://doi.org/10.1177/07342829241302860>
- Fidalgo, A. M., & Scalon, J. D. (2012). Uso dos métodos Mantel-Haenszel para a detecção do funcionamento diferencial dos itens e software relacionado [Using Mantel-Haenzel methods for detecting differential item functioning]. *Psicologia Reflexão e Crítica*, 25(1), 60-68. <https://doi.org/10.1590/S0102-797220120001000008>
- Fitriani, A., & Situmorang, D. D. B. (2023). The Differential Item Functioning (DIF) testing for the WOCC (Ways of Coping Checklist) instrument based on gender. *Heliyon*, 9, e14769. <https://doi.org/10.1016/j.heliyon.2023.e14769>
- Fruchart, E., & Rulence-Pâques, P. (2024). Development and validation of the Sports Grit Scale. *International Journal of Sport and Exercise Psychology*, 1-14. <https://doi.org/10.1080/1612197X.2024.2342029>
- Gao, F., Wei, Q., Dong, X., Gao, J., Lu, S., & Liu, Y. (2024). Developing and validating a domain-specific Grit Scale for college athletic students. *Scientific Reports*, 14, 11897. <https://doi.org/10.1038/s41598-024-62771-z>
- Gladstone, J. R., Morell, M., Yang, J. S., Ponnock, A., Turci Faust, L., & Wigfield, A. (2022). You can't compare if you don't prepare: Differential item functioning in measures of Grit, STEM self-regulation, and motivation. *The Journal of Experimental Education*, 91(3), 557-575. <https://doi.org/10.1080/00220973.2022.2062584>
- Gonzalez, O., Canning, J. R., Smyth, H., & MacKinnon, D. P. (2020). A psychometric evaluation of the Short Grit Scale: A closer look at its factor structure and scale functioning. *European Journal of Psychology Assessment*, 36, 646-657. <https://doi.org/10.1027/1015-5759/a000535>
- Guelmami, N., Chalghaf, N., Tannoubi, A., Puce, L., Azaiez, F., & Bragazzi, N. L. (2022). Initial development and psychometric evidence of Physical Education Grit Scale (PE-Grit). *Frontiers in Public Health*, 10, 818749. <https://doi.org/10.3389/fpubh.2022.818749>
- Gyamfi, A., & Acquaye, R. (2023). Parameters and models of Item Response Theory (IRT): A review of literature. *Acta Educationis Generalis*, 13(3), 68-78. <https://doi.org/10.2478/atd-2023-0022>
- Hasan, H. E. A., Munawar, K., & Khaiyom, J. H. A. (2022). Psychometric properties of developed and transadapted grit measures across cultures: A systematic review. *Current Psychology*, 41, 6894-6912. <https://doi.org/10.1007/s12144-020-01137-w>
- Hu, Z., Lin, L., Wang, Y., & Li, J. (2021). The integration of classical testing theory and item response theory. *Psychology*, 12(9), 1397-1409. <https://doi.org/10.4236/psych.2021.129088>
- Jafaripour, S., Tabatabaei, O., Salehi, H., & Dastjerdi, H. V. (2024). Applying IRT model to determine gender and discipline-based DIF and DDF: A study of the IAU English proficiency test. *International Journal of Language Testing*, 14(1), 56-74. <https://doi.org/10.22034/IJLT.2023.407117.1268>
- Jiang, L., Zhang, S., Li, X., & Luo, F. (2021). How grit influences high school students' academic performance and the mediation effect of academic self-efficacy and cognitive learning strategies. *Current Psychology*. <https://doi.org/10.1007/s12144-020-01306-x>
- Jie, W., Xinyi, W., & Tao, X. (2024). Reliability and validity of simple Chinese version of grit scale for elementary school students. *BMC Psychology*, 12, 304. <https://doi.org/10.1186/s40359-024-01809-3>
- Jones, E. A., Wind, S. A., Tsai, C. L., & Ge, Y. (2023). Comparing person-fit and traditional indices across careless response patterns in surveys. *Applied Psychological Measurement*, 47(5-6), 365-385. <https://doi.org/10.1177/01466216231194358>
- Kuruveettissery, S., Gupta, S., & Rajan, S. K. (2023). Development and psychometric validation of the Three-Dimensional Grit Scale. *Current Psychology*, 42, 5280-5289. <https://doi.org/10.1007/s12144-021-019-862-w>

- Li, J., Fang, M., Wang, W., Sun, G., & Cheng, Z. (2018). The influence of grit on life satisfaction: Self-esteem as a mediator. *Psychologica Belgica*, 58(1), 51-66. <https://doi.org/10.5334/pb.400>
- Li, C., & Yang, Y. (2023). Domain-general grit and domain-specific grit: Conceptual structures, measurement, and associations with the achievement of German as a foreign language. *International Review of Applied Linguistics in Language Teaching*, 62(4), 1513-1537. <https://doi.org/10.1515/iral-2022-0196>
- Lin, Y.-H., & Shih, C.-L. (2021). Analysis on mathematics Grit Scale and its gender-based DIF. *Asian Journal of Research in Education and Social Sciences*, 3(2), 73-79. <https://myjms.mohe.gov.my/index.php/ajress/article/view/14080>
- Linacre, J. M., & Wright, B. D. (2002). Construction of measures from many-facet data. *Journal of Applied Measurement*, 3(4), 486-512. <https://pubmed.ncbi.nlm.nih.gov/12486312>
- Marentes-Castillo, M., Zamarripa, J., & Castillo, I. (2019). Validation of the Grit Scale and the Treatment Self-Regulation Questionnaire (TSRQ) in the Mexican context. *Revista Latinoamericana de Psicología*, 51(1), 918. <https://doi.org/10.14349/rlp.2019.v51.n1.2>
- Matore, M. E. E. M., Mohamad, N., Khairani, A. Z., & Segar, T. N. (2023). Psychometric assessment of the short Grit Scale (Grit-S) using the Rasch Model in a Malaysian polytechnic students' sample. *Environment-Behavior Proceedings Journal*, 8(SI15), 127-132. <https://doi.org/10.21834/e-bpj.v8iSI15.5088>
- Midkiff, B., Langer, M., Demetriou, C., & Panter, A. (2017). Measuring grit among first-generation college students: A psychometric analysis. In L. A. Van der Ark, M. Wiberg, S. A. Culpepper, J. A. Douglas, & W. C. Wang (Eds.), *Quantitative psychology: IMPSS 2016* (Vol. 196, pp. 407-420). Springer. https://doi.org/10.1007/978-3-319-56294-0_35
- Monteiro, S., Postigo, Á., & Almeida, L. (2023). Escala Grit de Oviedo: Precisão e validade dos resultados da versão portuguesa [Oviedo Grit Scale accuracy and validity of results of the Portuguese version]. *Revista Iberoamericana de Diagnóstico y Evaluación Psicológica*, 67(1), 147-156. <https://doi.org/10.21865/RIDEP67.1.11>
- Morell, M., Yang, J. S., Gladstone, J. R., Turci Faust, L., Ponnock, A. R., Lim, H. J., & Wigfield, A. (2021). Grit: The long and short of it. *Journal of Educational Psychology*, 113(5), 1038-1058. <https://doi.org/10.1037/edu0000594>
- Muenks, K., Wigfield, A., Yang, J. S., & O'Neal, C. R. (2017). How true is Grit? Assessing its relations to high school and college students' personality characteristics, self-regulation, engagement, and achievement. *Journal of Educational Psychology*, 109(5), 599-620. <https://doi.org/10.1037/edu0000153>
- Nakano, T. C., Primi, R., & Nunes, C. H. S. S. (2015). Análise de itens e teoria de resposta ao item [Item analysis and item response theory]. In C. S. Hutz, D. R. Bandeira, & C. M. Trentini (Orgs.), *Psicometria* (pp. 97-124). Artmed.
- Neroni, J., Meijs, C., Kirschner, P. A., Kate, M. X., & Groot, R. H. M. (2022). Academic self-efficacy, self-esteem, and grit in higher online education: Consistency of interests predicts academic success. *Social Psychology of Education*, 25, 951-975. <https://doi.org/10.1007/s11218-022-09696-5>
- Noronha, A. P. P., & Almeida, L. S. (2022). A construção e estudos psicométricos da Escala de Avaliação da Garra: Versão internacional em Língua Portuguesa (EAGrit-LP) *Psicologia, Educação e Cultura*, 26(1), 8-23. <http://hdl.handle.net/10400.26/40365>
- Noronha, A. P. P., Reppold, C. T., & Almeida, L. S. (2024). Internal structure and invariance analysis of EAGrit-LP-R for Grit Assessment. *International Journal of Testing*, 24(2), 130-144. <https://doi.org/10.1080/15305058.2024.2312105>
- Pajestka, G., & Poraj-Weder, M. (2023). Persistence as a multidimensional construct: A new insight into the multidimensional persistence scale. *International Journal of Selection and Assessment*, 32(2), 195-209. <https://doi.org/10.1111/ijsa.12457>

- Park, H., Lee, K., & Shin, N. (2020). Development and validation of the Clinical Nurses Grit Scale (CN-Grit). *Journal of Korean Academy of Nursing Administration*, 26(1), 55-64. <https://doi.org/10.11111/jkana.2020.26.1.55>
- Peixoto, E. M., Honda, G. C., Gagnon, J., Nakano, T. C., Rocha, G. M. A., Zanini, D. S., & Balbinotti, M. A. A. (2019). Questionário de Fusão Cognitiva (CFQ): Novas evidências de validade e invariância transcultural. *Psico*, 50(1), 1-11. <https://doi.org/10.15448/1980-8623.2019.1.27851>
- Pendyala, A. G., & Vyas, M. (2023). Moving away from the WEIRD grit: A critical review. *Indian Journal of Positive Psychology*, 14(1), 67-71.
- Postigo, A., Cuesta, M., García-Cueto, E., Menéndez-Aller, A., Gonzáles-Nuevo, C., & Muniz, J. (2021). Grit assessment: Is one dimension enough?. *Journal of Personality Assessment*, 103(6), 786-796. <https://doi.org/10.1080/00223891.2020.1848853>
- Reysen, R., Reysen, M., Perry, P., & Knight, R. D. (2019). Not so soft skills: The importance of Grit to college student success. *Journal of College Orientation, Transition, and Retention*, 26(2). <https://doi.org/10.24926/jcotr.v26i2.2397>
- Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, 29(4), 347-363. <https://doi.org/10.1177/0734282911406661>
- Schimschal, S. E., Visentin, D., Kornhaber, R., & Cleary, M. (2022). Development of a scale to measure the psychological resources of grit in adults. *Nursing and Health Sciences*, 24(3), 752-763. <https://doi.org/10.1111/nhs.12973>
- Schmidt, F. T. C., Sudzina, F., & Botek, M. (2021). Psychometric assessment of the Short Grit Scale among Czech young adults. *Journal of Psychoeducational Assessment*, 39(4), 508-513. <https://doi.org/10.1177/0734282920974817>
- Sigmundsson, H., Haga, M., & Hermundsdottir, F. (2020). Passion, grit, and mindset in young adults: Exploring the relationship and gender differences. *New Ideas in Psychology*, 59, e100795. <https://doi.org/10.1016/j.newideapsych.2020.100795>
- Smith, E. V. (2004). Detecting and evaluating the impact of multidimensionality using the fit statistics and principal component analysis of residuals. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models, and applications* (pp. 575-600). JAM Press.
- Sordia, N. (2020). Psychometric properties of the Georgian version of the Grit Scale. *Prizren Social Science Journal*, 4(1), 8-13. <https://doi.org/10.32936>
- Sudina, E., Vernon, T., Foster, H., Del Villano, H., Hernandez, S., Beck, D., & Plonsky, L. (2020). Development and initial validation of the L2-Teacher Grit Scale. *TESOL Quarterly*, 55(1), 156-184. <https://doi.org/10.1002/tesq.581>
- Tan, C., Low, S. R., Chong, H. Y., Chong, S. L., Ong, A. W., Siah, P. C., Phang, S. F., Ong, Z. Q., Tan, W. H., Wong, J. A., & Lew, W. H. (2019). Exploratory and confirmatory factor analyses of the Short Grit Scale (Grit-S) for Malaysian undergraduate students. *Makara Human Behavior Studies in Asia*, 23(1), 27-33. <https://doi.org/10.7454/hubs.asia.2120519>
- Teimouri, Y., Sudina, E., & Plonsky, L. (2021). On domain-specific conceptualization and measurement of Grit in L2 learning. *Journal for the Psychology of Language Learning*, 3(2), 156-165. <https://doi.org/10.52598/jpll/3/2/10>
- Tynan, M. C. (2021). Deconstructing Grit's validity: The case for revising Grit measures and theory. In L. E. van Zyl, C. Olckers, & L. van der Vaart (Eds.), *Multidisciplinary Perspectives on Grit*. Springer, Cham. https://doi.org/10.1007/978-3-030-57389-8_8

- 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*, 35(4), 469-478. <https://doi.org/10.1027/1015-5759/a000424>
- Tyumeneva, Y., Kuzmina, J., & Kardanova, E. (2014). *IRT analysis and validation of the Grit Scale: A Russian investigation*. National Research University. https://www.researchgate.net/publication/314496371_IRT_Analysis_and_Validation_of_the_Grit_Scale_A_Russian_Investigation
- van Zyl, L. E., Heijenk, B., Klibert, J., Shanbkland, R., Verger, N. B., Rothmann, S., Cho, V., Feng, K., See-To, E. W. K., Roll, L. C., & van der Meij, L. (2022). Grit across nations: The cross-national equivalence of the Grit-O Scale. *Journal of Happiness Studies*, 23, 3179-3213. <https://doi.org/10.1007/s10902-022-00543-0>
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Mesa Press.
- Yiğiter, M. S., & Boduroğlu, E. (2024). Item response theory: A comprehensive review of studies with document analysis. *International Journal of Educational Studies and Policy*, 5(2), 119-138.
- Yupanqui-Lorenzo, D. E., Becerra-Herrera, W. A., Díaz-Leon, I., Sernaqué, M. A. C., & Olivera-Carhuaz, E. S. (2024). Short Grit Scale (Grit-S): New evidence based on CFA and MIRT models. *Revista Fuentes*, 26(1), 109-119. <https://doi.org/10.12795/revistafuentes.2024.23182>
- Zhang, T., Park, D., Tsukayama, E., Duckworth, A., & Luo, L. (2023). Sparking virtuous cycles: A longitudinal study of subjective well-being and grit during early adolescence. *Journal of Youth and Adolescence*. <https://doi.org/10.1007/s10964-023-01862-y>

Análise dos itens da Escala de Avaliação da Garra: Versão internacional em língua portuguesa

Resumo: A garra (grit) pode ser compreendida como um elemento da personalidade, sendo uma característica humana positiva e não cognitiva, composta por dois componentes: perseverança de esforço e consistência de interesses. Devido ao seu impacto social, profissional e educacional, a garra tornou-se um construto de grande interesse. Embora alguns instrumentos tenham sido desenvolvidos e adaptados para sua avaliação em diferentes contextos, a ausência de medidas em língua portuguesa levou ao desenvolvimento da Escala de Avaliação da Garra: Versão Internacional em Língua Portuguesa (EAGrit-LP), com base em amostras do Brasil e de Portugal. Apesar de estudos anteriores terem investigado algumas evidências de validade do instrumento, seus itens ainda não haviam sido avaliados individualmente. Assim, o objetivo do presente estudo foi realizar uma análise dos itens, utilizando a Teoria de Resposta ao Item. Para isso, uma amostra de 1.050 brasileiros e 656 portugueses respondeu à EAGrit-LP (com idades entre 17 e 71 anos; $M = 24,1$ anos; $DP = 9,2$). Os resultados mostraram que todos os itens apresentaram índices de infit e outfit adequados, e que a dificuldade dos itens foi, em geral, baixa. Seis itens apresentaram funcionamento diferencial do item (differential item functioning; DIF), com base no gênero e no país de origem dos participantes. Esses resultados podem ser utilizados para revisar a escala e orientar o desenvolvimento de itens mais desafiadores, bem como para a construção de tabelas normativas específicas para gênero e país de origem.

Palavras-chave: Psicologia positiva, Psicometria, Construção de testes, Avaliação psicológica.