Kenan Gao, Juan Zhang*, Yihui Wang, Chenggang Wu and Wei He

Development and validation of the L2 grit scale in collective cultural context (L2GSC): evidence based on machine learning methods

https://doi.org/10.1515/jccall-2024-0018
Received September 4, 2024; accepted December 13, 2024; published online March 4, 2025

Abstract: Prior research has demonstrated that in collectivist cultures (i.e., China and the Philippines), a Triarchic Model of Grit, encompassing perseverance of effort, consistency of interests, and adaptability to situations, can be more fitting. This study explores the concept of second language (L2) grit within a collectivist context, investigating its association with positive psychological constructs in L2 learning among Chinese college students studying English as a foreign language (n = 459). In this pursuit, a specialized grit scale tailored for language learning, the L2 Grit Scale in Collective Cultural Context (L2GSC), was developed and validated to assess the domain-specific grit of L2 learners. The findings based on hierarchical regression and machine learning methods revealed that L2GSC positively predicted learners' enjoyment and engagement in language learning, exceeding the effects found with triarchic model of grit scale, suggesting that L2GSC is suited to collectivist cultures.

Keywords: L2 grit; triarchic model of grit; second language acquisition; collective culture; machine learning

1 Introduction

Currently, there's a rising trend among second language acquisition (SLA) researchers recognizing the significance of positive psychology (PP) constructs (Padilla

***Corresponding author: Juan Zhang**, Faculty of Education, University of Macau, Macao, China, E-mail: juanzhang@um.edu.mo. https://orcid.org/0000-0002-7052-1093
Kenan Gao and Yihui Wang, Faculty of Education, University of Macau, Macao, China, E-mail: yc37104@connect.um.edu.mo (K. Gao), yc17105@connect.um.edu.mo (Y. Wang). https://orcid.org/0009-0004-7476-9514 (K. Gao). https://orcid.org/0000-0002-3221-3313 (Y. Wang)
Chenggang Wu, School of Education, Shanghai International Studies University, Shanghai, China, E-mail: chenggangwu@shisu.edu.cn. https://orcid.org/0000-0003-3837-3841
Wei He, Department of Engineering Science, Macau University of Science and Technology, Macao, China, E-mail: vicchoo@outlook.com. https://orcid.org/0000-0001-7786-2715

et al. 2020). Grit, as a distinctive instance of PP variables, plays a significant role in PP by contributing to happiness and life satisfaction (Khan and Khan 2017). It is initially conceptualized as a fundamental and enduring trait that manifests in a person's unwavering commitment and effort toward achieving long-term goals, regardless of any obstacles, challenges, or slow advancements (Duckworth et al. 2007). The authors argue that there are two lower-order constructs: perseverance of effort (PE) and consistency of interest (CI). PE is characterized by an individual's capacity to uphold dedication and continue to exert effort when faced with challenges. CI, on the other hand, pertains to the sustained focus and enthusiasm for a particular interest or goal over an extended period of time. According to Von Culin et al. (2014), grit is a relatively stable characteristic that persists over various circumstances and timeframes. This implies that individuals with higher levels of grit tend to succeed through a persistent and engaged effort, regardless of potential setbacks or challenges. In addition, research indicates that grit, potentially more than innate talent, contributes to student success by offering added predictive power for achievement beyond natural abilities (Duckworth and Quinn 2009; Duckworth et al. 2007).

Building on this concept, Datu et al. (2017) introduced the Triarchic Model of Grit Scale (TMGS), designing to measure an individual's propensity to remain determined in pursuing long-term goals within a collectivist/group-oriented context. The reason why developing a TMGS is that individuals in a collectivist/group-oriented or an individualist-oriented context are different. According to Littlewood (1999), in a collectivist context, individuals perceive themselves as integral to their group, prioritizing the group's perspectives, needs, and objectives over individual distinction. Individualism is characterized by an emphasis on the "I", while collectivism emphasizes the "We" (Littlewood 1999). Therefore, this model emphasizes not just the importance of PE and CI in chasing long-term goals, but also the importance of adaptability to situations (AS), which means being adaptable and able to adjust one's actions and strategies to align with varying social, situational, and environmental factors. As explained by Datu and Zhang (2021), in societies where people value a context-sensitive self (Suh 2007), the capacity to adapt is often seen as a precursor to achieving beneficial outcomes.

While general grit has been widely investigated, some researchers have posited that grit should be viewed as a trait that manifests differently across specific fields rather than a uniform, general quality (Duckworth and Quinn 2009; Teimouri et al. 2020). For example, in SLA, grit is posited as essential for sustained learning efforts in second or foreign language (henceforth L2) learners (Sudina et al. 2021). This has led to a rise in investigations into how grit affects language learning, revealing its importance in shaping a learner's motivation (Khajavy et al. 2021), willingness to communicate (Lan et al. 2021), emotional well-being (Sudina and Plonsky 2021b), and overall language proficiency (Fathi and Hejazi 2024; Sudina and Plonsky 2021b).

Nonetheless, Teimouri et al. (2020) evidenced that evaluating personality through measures specific to the language domain yields a more accurate assessment of students' personality traits and behavioral characteristics in L2 contexts, compared to broad, out-of-context personality assessments. This approach underlines that grit levels can differ significantly across various life contexts (Cormier et al. 2019). To be specific, the domain-specific view of grit focuses on its varying degrees of intensity within distinct achievement areas, leading to a more precise and locally consistent measurement of traits (Cormier et al. 2019). Therefore, the effectiveness of a general grit scale is questionable in specific fields (Cormier et al. 2019; Li and Yang 2023).

Compared with domain-general grit, domain-specific grit (i.e., L2 grit) has increased power in predicting language academic achievement in L2 (Teimouri et al. 2020). L2 grit is defined as the specific application of passion and perseverance to the learning of an L2 (Teimouri et al. 2020). Prior research suggests a correlation between L2 grit and an extensive array of beneficial outcomes in L2 learning, showing that L2 grit is associated with lower foreign language anxiety (Liu et al. 2023a), higher foreign language enjoyment (Fathi and Hejazi 2024), increased growth mindset (Fathi et al. 2024), strengthened self-efficacy beliefs (Fathi et al. 2024), stronger academic engagement (Sadoughi and Hejazi 2023), and better foreign language performance (Fathi et al. 2024). These relationships can be explained through both theoretical and empirical perspectives.

Both self-determination theory (SDT, Deci and Ryan 1985) and broaden-and-build theory (BBT, Fredrickson 2003) focus on individual growth. SDT focuses on how the satisfaction of motivational needs drives personal growth, while BBT focuses on how positive emotions promote adaptation and resilience by expanding cognition and accumulating resources. The combination of the two provides a complete perspective for understanding how individuals can achieve lasting personal development with the support of intrinsic motivation and positive emotions.

SDT provides a robust framework for understanding the relationship between grit, enjoyment and engagement, particularly by explaining how perseverance in long-term goal pursuit fosters positive emotional experiences when autonomy, competence, and relatedness needs are met (Deci and Ryan 1985; Niemiec and Ryan 2009; Ryan and Deci 2000). Individuals with high grit typically pursue their chosen long-term goals voluntarily (autonomy), which enables them to maintain effort despite challenges, build skills, and gain experience (competence), often accompanied by social support from others (relatedness). Through persistent effort, achievement, and recognition, such individuals are likely to be more focused, active, thus more engaged in the process of achieving goals and experiencing enhanced enjoyment throughout the process.

According to the BBT (Fredrickson 2003), positive emotions broaden individuals' cognitive and behavioral capacities, helping them build long-term resources such as

psychological resilience, social connections, and skills, which can be beneficial in coping with challenges in the future. This theory provides a framework for understanding the relations between L2 enjoyment, L2 grit, and L2 engagement. On the one hand, in the context of language learning, L2 enjoyment, as a "positive learning-facilitative achievement emotion," (Dewaele and Li 2021, p. 926) is closely associated with motivation, engagement, and L2 performance. L2 grit enables individuals to persist through challenges, which increases the likelihood of experiencing enjoyment as they progress toward their goals. As enjoyment increases, individuals may further broaden their experiences and invest more deeply in their pursuits, which strengthens grit over time. On the other hand, L2 grit as a personal resource enables individuals to broaden their experiences and cognitive responses, leading to sustained engagement in meaningful tasks. As people remain engaged, they build further emotional and psychological resources (e.g., enthusiasm and motivation), reinforcing their grit and ability to maintain focus and effort over time. As a result, the relationships among language L2 enjoyment, L2 engagement, and L2 grit are complementary, with active engagement enhancing learners' grit, thereby fostering long-term success in language learning.

Accordingly, Liu et al. (2023a) highlight that emotional and instrumental support from teachers is crucial in fostering L2 grit, which significantly helps reduce foreign language anxiety. This underscores the need for educators to focus on both instructional content and providing emotional and practical support to improve student outcomes (Liu et al. 2023a). For instance, Zhao (2023) reviewed nine studies on the relationship between grit and foreign language enjoyment, concluding that these two variables are positively correlated. Thus, L2 enjoyment was used as a predicted outcome of L2 grit in this study. Similarly, Liu et al. (2023b) suggest that learners experiencing burnout struggle with engagement unless they build resilience and motivation. Resilience helps individuals stay committed despite setbacks (Eskreis-Winkler et al. 2014), while grit, emphasizing perseverance and overcoming obstacles for long-term goals (Duckworth et al. 2007), correlates positively with engagement and performance among EFL learners (Eren and Rakıcıoğlu-Söylemez 2023). Therefore, empirical studies demonstrated that grit supports enhanced student engagement, and L2 enjoyment was again employed as a predicted outcome of L2 grit in this study.

To date, two prominent L2-related grit scales have been developed, namely L2 grit (Teimouri et al. 2020) and L2 grit scale (Alamer 2021), both rooted in the seminal work of Duckworth et al. (2007). Teimouri et al. (2020)'s 9-item L2 grit scale, situated in the Iranian context, confirmed the existence of two significant components (i.e., PE and CI). In addition, Alamer (2021) devised and validated a 12-item L2 grit scale in the Saudi Arabian context based on the Grit–S scale developed by Duckworth et al.

(2007), confirming that grit as a single construct with two specific scales, consistent with the initial conceptualization.

While prior studies (e.g., Liu et al. 2022) have confirmed the reliability and validity of the L2 grit scales developed by Teimouri et al. (2020) and Alamer (2021), the results have varied across different cultural and educational contexts. Specifically, on the one hand, by using the L2 grit scale (Teimouri et al. 2020), Sudina et al. (2021) posited that PE exclusively has a significant correlation with L2 achievement across different languages and contexts. Conversely, Sudina and Plonsky (2021a) highlighted that CI exhibits greater predictive and incremental validity to L2 achievement than PE. Adding complexity to this debate, Cheng (2021) found that both PE and CI facets notably contribute to predicting the willingness to communicate in L2 among English majors in Taiwan, merely PE had a more pronounced impact than CI. Similarly, Wu et al. (2024) recently found that PE rather than CI could predict language learning engagement. On the other hand, in the context of the Middle East, Alamer (2021)'s L2 grit scale also has mixed outcomes. For example, Khajavy et al. (2021) reported no correlation between grit (encompassing both PE and CI) and L2 achievement among Iranian university students. In a subsequent study, Khajavy and Aghaee (2024) indicated that PE alone was a significant indicator of L2 success. Therefore, it becomes evident that the construct of grit in the realm of language learning is multifaceted and its influence on L2 achievement is not universal. The variance in findings underscores the importance of considering cultural and contextual factors when assessing the role of grit in language acquisition. While some components of grit such as PE may stand out as key predictors in certain environments, the overall picture is one of a nuanced trait interacting with diverse educational settings and learner backgrounds. This suggests that educators and researchers should adopt a more localized approach to understanding and applying the concept of grit in L2 learning, recognizing that what may be a significant factor in one context may not hold the same level of influence in another.

Given the nascent stage of promoting grit in PP, more empirical research is needed to refine its definition and measurement in L2 learning. Notably, no research has explored how L2 grit develops and operates from a collectivist/group-oriented perspective, as existing L2-related grit scales were primarily developed and tested from an individualist-oriented perspective (Alamer 2021; Teimouri et al. 2020). Considering that people in East Asian countries exhibit a stronger tendency towards collectivism compared to those in Western countries (Littlewood 1999), and a substantial amount of L2 grit research is carried out in such environments (e.g., China, Japan, and Korea) (Zhao and Wang 2023), it raises the question of whether an L2 grit scale, if developed and validated within a collective society, might yield greater applicability and relevance in such settings. In other words, it requires the necessity to know the potential effectiveness of an L2 grit scale developed and validated within

a collective society. To this end, this study aims to critically examine the core issues in L2 grit research. It highlights the oversight of collectivist cultural backgrounds in L2 grit research and the need for more culturally nuanced instruments to measure the personality traits of L2 learners from collective cultural backgrounds.

Machine learning (ML) is increasingly utilized in psychological and linguistic research to enhance the accuracy, efficiency, and predictive power of psychometric and language learning assessments (Lin 2021; Shatte et al. 2019). ML algorithms, such as random forests, can reveal subtle relationships between variables that traditional statistical methods may overlook (Shatte et al. 2019). For instance, a review points out ML has been successfully applied in applied linguistic studies and enjoyed a promising future (Lin 2021). Unlike traditional linear regression with its rigid assumptions of linearity, ML demonstrates significant potential for future studies (Lin 2021). In addition, prediction with ML can lead to a deeper understanding of behavior (Yarkoni and Westfall 2017). In our study, we aim to apply ML in an innovative way to explore L2 grit scale in collective cultural context (L2GSC) to capture L2 learning-related variables that conventional methods might overlook, providing more precise predictions as noted above. Furthermore, many ML algorithms make minimal assumptions about data distribution, for example, allowing for effective handling of non-normally distributed data. Therefore, our approach is informed by both theoretical considerations and empirical evidence, leveraging ML to uncover insights that contribute to the advancement of L2 learning research.

Grounded in the foundational research on the construct of grit (Alamer 2021; Datu et al. 2016, 2017; Duckworth et al. 2007; Duckworth and Quinn 2009; Teimouri et al. 2020), the study aims to conceptualize, develop, and empirically validate a L2GSC. This endeavor is informed by the TMGS proposed by Datu et al. (2017). Within this framework, L2 grit can be posited as a nuanced trait characterized by the PE, CI, and AS within SLA. In addition to a new model of grit specific to L2 contexts, its relevance and potential impact on the behaviors of L2 learners will also be examined, considering that L2 grit may imply different meanings in different cultural contexts (Datu et al. 2017).

2 Phase 1 item pool of L2GSC

2.1 Participants

Based on the concept of L2 grit, the interviewees were selected from college students who had faced language learning challenges but subsequently strived to overcome them and improve in foreign language learning. Through coordination with instructors from two universities in China, 30 college students from different majors

(i.e., Humanities, Social Science, and Medicine) studying English as an L2 were initially identified as potential candidates by convenient sampling. After 10 interviews, the researchers observed that the data reached saturation. To ensure thoroughness, an additional 6 interviews were conducted, totaling 16 interviews. The final sample included 13 undergraduates, 2 master's students, and 1 doctoral student; 9 males and 7 females, with an average age of 19.81 years (SD = 1.97).

2.2 Interview procedure

Four research assistants engaged in online individual interviews with participants via Tencent Meeting. Before the interview, informed consent was obtained from all the participants, acknowledging that the interviews were unrelated to their educational institution and were solely for academic inquiry. All 16 participants consented to audio recording throughout the interviews. The structured interview protocol with six questions was designed to mitigate variances arising from different interviewers. The two core inquiries focused on the participants recounting their most significant setbacks in learning English from childhood to adulthood and their strategies to overcome these challenges and enhance their proficiency. Supplementary questions were crafted to facilitate recollecting past experiences, such as assessing the individual's role in the adaptation process. The interview language was Putonghua. Each interview lasted approximately 25 min.

2.3 Interview analysis and results

This study followed a systematic approach to organize the original audio recordings: (1) Transcription: Each interview was transcribed into text promptly after completion to ensure data accuracy. (2) Content Analysis: The transcript text was analyzed using a step-by-step content-coding. Text materials were first categorized based on content. Ambiguous or unclear information was independently coded and then reviewed for secondary step, incorporating it into the initial categories after determining its meaning. Following these steps, the data was organized into three categories: (1) Diligence and Persistence (e.g., I am a diligent English learner). (2) Enthusiasm and Interest (e.g., I continue to maintain my enthusiasm for learning English). (3) Adaptation and Adjustment (e.g., I adjust my English learning strategy in a timely manner).

The validity of these classifications was verified through feedback from interviewees and an independent third party (a professor in Psychology) with expertise in qualitative research. The data revealed that the three identified categories aligned well with Datu et al. (2017)'s dimensions, leading to the adoption of category names of Datu et al. (2017). In developing the scale, questions were refined, with similar yet distinct items retained for the initial test, such as "As long as I decide to

learn English well, I will firmly achieve my goals despite any difficulties," and "Challenges in learning English motivate me to work harder." Ultimately, 15 items were included in the initial test.

3 Phase 2 initial test of L2GSC

The primary objective of phase 2 was to assess its factor structure and score reliabilities. This involved employing exploratory factor analysis (EFA) and conducting reliability analyses using Cronbach's alpha.

3.1 Participants

Data were gathered through online surveys, employing convenience sampling to recruit 222 participants from chinese universities. Among these participants, 177 (79.7 %) were female, and 45 (20.3 %) were male. The predominant age group was between 18 and 25, covering 99.1 % of the participants, while 0.9 % were over 25. Notably, 209 (94.1 %) participants were undergraduate students, with 10 more master's students and 3 doctorate students.

3.2 Procedure and statistical analysis

Intraclass correlation coefficients (ICC) was used to assess inter-rater agreement. EFA was performed utilizing principal axis factoring and promax rotation in 26th SPSS software to identify items that accurately represented the intended constructs. Internal consistency was assessed using Cronbach's alpha coefficient, along with descriptive statistics and item-total correlations, to determine the reliability of the items.

3.3 Results

3.3.1 The initial scale

Inter-rater agreement was gauged using ICC (two-way mixed model, mean of 5 raters, absolute agreement). Following the removal of two items due to limited content relevance, the panel determined that the ICC for the 13-item scale was 0.87. The initial L2GSC employed a 5-point Likert-type scale, ranging from 1 (Not like me at all) to 5 (Very much like me). Specific items are delineated in Table 1.

Table 1: Results of the principal component analysis of the initial 13-item scale (only Chinese items were used in this study).

	Perseverance	Consistency	Adaptability
1. 就学习英语而言,我是一个努力学习的人。 (When it comes to English, I am a hard-working learner.)	0.744		
2. 只要我决定学好英语,无论遇到什么困难,我都会坚定地实现我的目标。 (Now that I have decided to learn English, nothing can prevent me from reaching this goal.)	0.703		
3. 我是一个勤奋的英语学习者。 (I am a diligent English language learner.)	0.873		
4. 无论遭遇任何挫折,我都坚决不放弃英语学习。 (I will not allow anything to stop me from my progress in learning English.)	0.726		
5. 我会花很多时间和精力来提高我的英语薄弱点。 (I put much time and effort into improving my English language weakness.)	0.608		
6. 我现在对学习英语的兴趣不如以前那么浓厚。 (I am not as interested in learning English as I used to be.)		0.787	
7. 我曾经对语言学习很感兴趣,但最近我的兴趣似乎有所减退。 (I was obsessed with learning English in the past but have lost interest recently.)		0.753	
8. 我对英语学习的兴趣一直在变。 (My interests in learning English changes from year to year.)		0.655	
9. 我已经失去了学习英语的兴趣。 (I think I have lost my interest in learning English.)		0.586	
10. 我愿意探索英语学习的新方法。 (I appreciate new English language learning opportunities.)			0.777
11. 适时调整我的学习策略对于实现我的长期语言目标至关重要。 (Adapting my study methods is vital to achieving my long-term English proficiency goals.)			0.912
12. 我能够适应英语学习中的不同环境或挑战。 (I am able to adjust to new English learning environments or challenges.)			0.493
13. 我始终有动力去提高我的语言学习技能。 (I am constantly driven to enhance my English learning competencies.)			0.556

3.3.2 Factor structure and reliability

Initially, the mean, standard deviation, and correlations between individual items and the total score for the initial L2GSC were calculated to determine the items that should be retained. The results indicated that most items exhibited comparable means, appropriate dispersions, and item-total correlations. Importantly, none of the items had item-total correlation values dropped below 0.4. Therefore, no items were eliminated during this phase of analysis (Field 2024).

Following this, a factor analysis was performed on the remaining set of 13 items in the L2GSC. The Kaiser-Meyer-Olkin (KMO) value, registering at 0.79, suggests that the data is appropriate for factor analysis according to the criteria set by Hutcheson and Sofroniou (1999). An examination of the Eigenvalues disclosed the presence of three factors in the data, with values of 4.37, 1.99, and 1.18, respectively, consistent with the theoretical underpinnings of L2GSC. Items on PE, CI, and AS in language learning constituted the first, second, and third factors, respectively. Collectively, these factors explained 58.01 % of the L2 learning grit score variance. All factors exhibited satisfactory loadings on their respective components. Subsequently, Cronbach's alpha reliability coefficients for each subscale was computed, resulting in the following values: α in perseverance = 0.81; α in consistency = 0.65; α in adaptability = 0.77. These values point to robust internal consistencies, ranging from acceptable to good (Black et al. 2010).

4 Phase 3 refining the L2GSC

Phase 3 focused on reinforcing the structural validity of the L2GSC via Confirmatory Factor Analysis (CFA). Subsequent evaluations of construct validity were conducted through correlational analysis, while criterion-related validity was assessed using hierarchical regression and ML techniques, with criteria including foreign language engagement and enjoyment.

4.1 Participants

In Phase 3, a new convenient sample of university students from mainland China, totaling 237 individuals, was utilized. Among these participants, 180 (75.9 %) were female, and 57 (24.1 %) were male. The predominant age bracket was 18–25, including 99.2 % of the participants, with only 0.8 % being over 25. Importantly, the majority of individuals (n = 234) were currently enrolled in their undergraduate studies, comprising 98.7 %, with 2 master's students and 1 doctorate student.

4.2 Statistical analysis

4.2.1 Confirmatory factor analysis (CFA)

The goodness of fit for the 13-item version of L2GSC was assessed using the maximum likelihood estimation method in CFA, employing the 23rd version of AMOS.

4.2.2 Correlational analysis

To gain preliminary insights into the nomological network of L2GS, which encompasses the interconnected relationships between L2 learning grit and other related psychological variables, the associations between L2 learning grit subscales and general grit, conscientiousness, and language growth mindset were examined. This analysis aimed to assess the construct validity of L2 grit as a distinct personality trait by examining its associations with other related traits.

4.2.3 Hierarchical regression analysis

Moreover, to assess the adaptability of L2GSC to the L2 learning environment, hierarchical regression analyses were conducted using SPSS 26th. This involved controlling for demographic variables and the general grit, following the approach used by Datu et al. (2017). This approach examined the predictive power of the three dimensions of L2 grit on relevant L2 variables, including L2 learning engagement and enjoyment.

4.2.4 Machine learning analysis

ML was also used to examine the adaptability of L2GSC to the L2 learning environment. ML excels in processing and analysing large volumes of data, enhancing the accuracy of predictions. It is capable of learning from language-related data and forecasting outcomes (Rezaii et al. 2019). This capability enables researchers and educators to more effectively identify and comprehend the patterns and trends within the language learning process. In this study, three models were employed: Random Forest (RF) and Gradient boosted decision trees (GBDT), which includes XGBoost and LightGBM.

Random Forest, introduced by Breiman (2001), is an ensemble technique that builds multiple decision trees to create a robust model for prediction. The algorithm is known for its simplicity, efficiency, and ability to handle large datasets with a high dimensionality. It operates by constructing numerous trees using bootstrap samples of the training data and making predictions by averaging the outputs of these trees.

This process enhances the model's generalization capability, making RF resistant to overfitting. The RF model's predictive performance can be represented as:

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \tag{1}$$

where T_i is the prediction of the ith tree and N is the number of trees.

GBDT is an ensemble technique that builds trees sequentially, with each tree learning from the errors of its predecessors. XGBoost (Chen and Guestrin 2016) and LightGBM (Ke et al. 2017) are two significant advancements in this domain, offering efficiency and scalability.

XGBoost employs Newton's method for optimization, utilizing the second-order Taylor expansion of the loss function for more accurate convergence. It introduces a regularization term to its objective function to control model complexity and reduce overfitting:

$$Obj(t) \approx \sum_{i=1}^{n}\left[l\left(y_i, \widehat{y_i^{(t-1)}}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right] + \Omega(f_t) \tag{2}$$

where l is the loss function, g_i and h_i are the first and second derivatives of the loss function, respectively, f_t is the output of the tth tree, and $\Omega(f_t)$ is the regularization term.

LightGBM focuses on reducing training time without sacrificing accuracy. It introduces Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB) to handle large datasets efficiently. GOSS keeps instances with large gradients, while EFB bundles mutually exclusive features, reducing complexity. Unlike XGBoost, LightGBM grows trees leaf-wise, selecting the split that results in the largest decrease in loss:

$$\text{Gain} = \text{Loss}_{\text{before}} - \text{Loss}_{\text{after}} \tag{3}$$

In conclusion, RF offers a robust and straightforward approach to ensemble learning, while XGBoost and LightGBM extend the capabilities of GBDT by introducing sophisticated optimization techniques and efficiency improvements. These algorithms provide powerful tools for tackling complex predictive tasks in ML.

The evaluation metrics included accuracy, precision, recall, and the F1-score. These metrics provided a holistic view of the models' performance, catering to different aspects of predictive accuracy and the balance between precision and recall. Accuracy measures the proportion of correct predictions (both true positives and true negatives) among all cases, calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{4}$$

Where TP (True Positives) are correctly predicted positive cases; TN (True Negatives) are correctly predicted negative cases; FP (False Positives) are negative cases incorrectly labeled as positive; FN (False Negatives) are positive cases incorrectly labeled as negative.

F1-Score is the harmonic mean of precision and recall, focusing on the balance between them, defined as:

$$\mathrm{F1-Score} = 2 \cdot \frac{\mathrm{preclsion - recall}}{\mathrm{preclision + recall}} \tag{5}$$

In equations for precision and recall:

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP + FP}} \tag{6}$$

$$\mathrm{Recall} = \frac{\mathrm{TP}}{\mathrm{TP + FN}} \tag{7}$$

Precision is the ratio of correctly predicted positive cases to all predicted positives. Recall is the ratio of correctly predicted positives to actual positives.

4.3 Measures

4.3.1 General grit

The Chinese adaptation of the TMGS (Datu and Zhang 2021), serves as a tool for evaluating an individual's general grit. This scale encompasses three key dimensions: perseverance, passion, and adaptability. Utilizing a 5-point Likert scale, respondents rate the extent to which each statement aligns with their personal experiences, with 1 indicating "not at all like me" and 5 indicating "very much like me". In the present study, the reliability of the scale was confirmed with a Cronbach's alpha coefficient of 0.70.

4.3.1.1 Conscientiousness

The measurement of conscientiousness is a subscale of Big Five personality traits. The Chinese adaption of the NEO Five-Factor Inventory (NEOFFI) was employed (Jiang 2020). The original scale consists of 60 items, and for the purposes of this study, the 12-item conscientiousness subscale was extracted. Each item is rated on a 5-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree). In this study, the Cronbach's alpha of conscientiousness was 0.61.

4.3.2 Growth mindset

Students' growth mindset toward their English learning abilities was assessed using a six-item scale adapted from Yao et al. (2021), with choices ranging from 1 (strongly disagree) to 6 (strongly agree). This scale comprises six questions, divided into three growth-mindset items and three fixed-mindset items. In this study, three growth-mindset items were used. The reliability of the scale was confirmed with a Cronbach's alpha of 0.82.

4.3.3 Foreign language engagement

The Foreign Language Engagement scale, formulated by Li et al. (2024), includes four sub-dimensions: cognitive engagement, emotional engagement, behavioral engagement, and social engagement. This 15-item scale uses a 5-point rating system (1 = Completely disagree, 5 = Completely agree) and was initially created and validated within a rural Chinese setting. In the present study, this scale was utilized to assess individuals' degrees of engagement in the process of learning an L2. The scale exhibited strong reliability (Cronbach's Alpha = 0.86).

4.3.4 Foreign language enjoyment

In this study, the 11-item Chinese Version of the Foreign Language Enjoyment Scale (CFLES) was used, which was validated by Li et al. (2018), to assess the enjoyment of L2 learning. This scale consists of three sub-dimensions: FLE-Private, FLE-Teacher, and FLE-Atmosphere. Its objective is to determine whether individuals encounter a supportive, enjoyable, and personally growth-oriented learning environment while studying foreign languages in China. The CFLES employs a conventional 5-point Likert scale, ranging from "1 represents strongly disagree" to "5 represents strongly agree". The reliability of CFLES was high, with a reported Cronbach's alpha coefficient of 0.91.

4.4 Results

4.4.1 CFA results

In the proposed measurement model, L2GSC was conceptualized with three latent factors, each representing one of the three sub-dimensions: perseverance, consistency, and adaptability. These latent factors are indicated by 5, 4, and 4 observed indicators, respectively.

To evaluate the model fit, established criteria for acceptable fit were applied (Schermelleh-Engel et al. 2003): 1) a non-significant χ^2; 2) Comparative Fit Index (CFI), Goodness of Fit Index (GFI), and Tucker-Lewis Index (TLI) exceeding 0.90; and 3) Root

Mean Square Error of Approximation (RMSEA) below 0.08. The initial CFA revealed suboptimal fit indices: χ^2 = 220.01, df = 62, p < 0.001, CFI = 0.89, GFI = 0.88, TLI = 0.89, and RMSEA = 0.10 (CI 90 % = 0.089–0.12). A scrutiny of the modification indices (MI) suggested that the inclusion of Item 9 in the consistency of interests was detrimental to the model fit (MI = 35 > 12.0). Following the methodological framework presented by Steger et al. (2006), Item 9 was omitted, leading to another CFA with the revised 12-item version of the L2GSC. This revision yielded improved and acceptable fit indices: χ^2 = 148.25, df = 51, p < 0.001, CFI = 0.93, GFI = 0.91, TLI = 0.93, and RMSEA = 0.09 (CI 90 % = 0.073–0.107) (Figure 1).

Following refinement, the reliability of the enhanced L2GSC is as follows: the overall scale exhibits a Cronbach's alpha of 0.87, and the alphas for the sub-

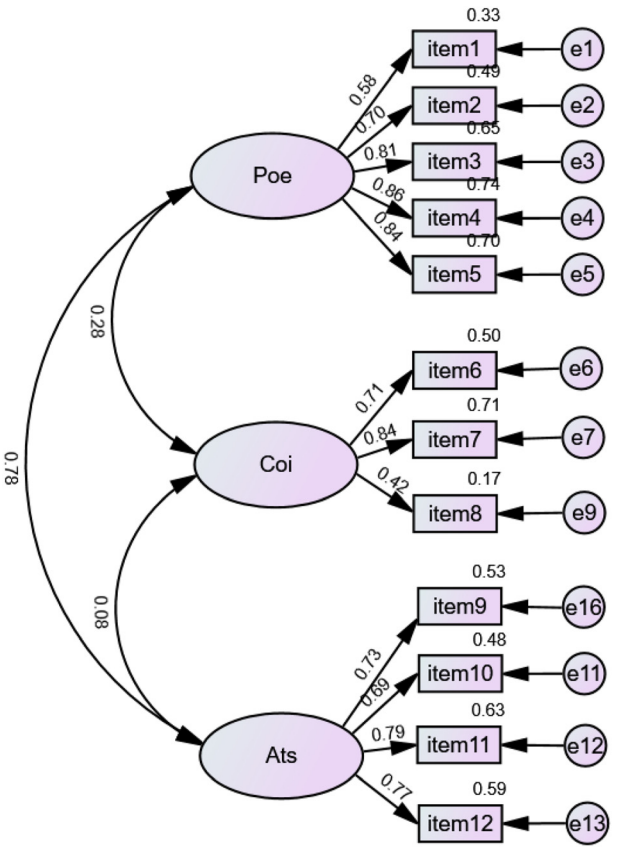

Figure 1: Outcome of confirmatory factor analysis in phase 3.

dimensions are reported as follows: α for perseverance is 0.87, α for consistency is 0.66, and α for adaptability is 0.83.

The correlation analyses demonstrated that PE, along with AS, exhibited positive associations with PE and AS of general grit, growth mindset, and conscientiousness. This reflected that PE and AS have strong correlations with other relevant psychological traits, indicating that these subscales can effectively measure relevant psychological traits in L2 learning. This supported the construct validity of L2 grit as an independent trait. However, consistency was not observed to correlate with adaptability in TMGS, or growth mindset. This may indicate that the subscale of consistency of interest may not be as closely related to relevant psychological variables in L2 grit as the other two subscales. Detailed results of descriptive statistics, reliability estimates, and correlational coefficients for the study variables are presented in Table 2.

4.4.2 Hierarchical regression results

Before the hierarchical regression analyses, descriptive statistics, reliability indices, and correlations for L2GSC dimensions, general grit (TMGS), L2 engagement, and L2 enjoyment were computed to ensure data quality, confirm the reliability of the measurement instruments, and identify preliminary relationships among the variables (Table 3). The outcomes of hierarchical regression analyses demonstrated that L2 grit had a strong positive effect in both models, that is, it was an important predictor of L2 engagement and enjoyment. This connection persisted even when considering demographic characteristics and trait grit. In addition, consistency and adaptability also show significant positive effects in Model 1. The introduction of L2 grit dimensions into the regression models highlighted additional variance (ranging from 4.7 % to 29 %) in the mentioned outcomes, emphasizing the substantial contribution of L2 grit in predicting L2 learning engagement and enjoyment (refer to Table 4).

4.4.3 Machine learning results

The analysis initiated a classification task aiming to predict outcomes from a dataset with two distinct targets: L2 enjoyment and L2 engagement. The predictive variables for these objectives are derived from both the three sub-dimensional scores and the overall scores of L2GSC and TMGS. This preprocessing step was crucial for converting the research aim (examine the adaptability of L2GSC to the L2 learning environment) into a multiclass classification scenario. The dataset was split 80–20 into training and testing sets for thorough model evaluation. The models were implemented in Python using libraries for Random Forest, XGBoost, and LightGBM.

Table 2: Descriptive statistics, reliability coefficients, and correlations of L2GSC with general grit, growth mindset, and conscientiousness.

	Mean	Std. Deviation	α	1	2	3	4	5	6	7	8
1. Perseverance	3.22	0.71	0.87	1							
2. Consistency	2.88	0.69	0.66	0.356**	1						
3. Adaptability	3.42	0.65	0.83	0.658**	0.189**	1					
4. gPerseverance	3.55	0.74	0.79	0.474**	0.300**	0.496**	1				
5. gConsistency	2.70	0.76	0.79	−0.090	0.172**	−0.239**	−0.419**	1			
6. gAdaptability	3.46	0.67	0.88	0.352**	0.031	0.444**	0.589**	−0.502**	1		
7. Growth mindset	4.00	0.99	0.84	0.468**	0.103	0.503**	0.501**	−0.340**	0.485**	1	
8. Conscientiousness	3.23	0.37	0.61	0.292**	0.269**	0.324**	0.455**	0.066	0.300**	0.165*	1

The gPerseverance, gConsistency and gAdaptability refer to corresponding constructs in TMGS respectively. *$p < 0.05$, **$p < 0.01$.

Table 3: Descriptive statistics, reliability coefficients, and correlations of L2GSC with L2 engagement, L2 enjoyment, and general grit.

	Mean	Std. Deviation	α	1	2	3	4	5	6	7	8	9	10
1. Perseverance	3.22	0.71	0.87	1									
2. Consistency	2.88	0.69	0.66	0.302**	1								
3. Adaptability	3.42	0.65	0.83	0.658**	0.138*	1							
4. L2 enjoyment	3.34	0.67	0.91	0.480**	0.220**	0.504**	1						
5. Cognitive engagement	3.32	0.81	0.90	0.451**	0.239**	0.507**	0.585**	1					
6. Emotional engagement	3.13	0.58	0.61	0.451**	0.370**	0.401**	0.443**	0.603**	1				
7. Behavioral engagement	3.28	0.77	0.90	0.473**	0.262**	0.453**	0.561**	0.787**	0.587**	1			
8. Social engagement	2.83	0.79	0.81	0.050	0.300**	-0.021	-0.076	-0.154*	0.092	-0.129*	1		
9. L2 engagement	3.14	0.50	0.86	0.514**	0.423**	0.487**	0.550**	0.819**	0.796**	0.820**	0.311**	1	
10. General grit	3.24	0.36	0.83	0.479**	0.299**	0.446**	0.528**	0.509**	0.422**	0.538**	0.109	0.578**	1

*$p < 0.05$, **$p < 0.01$.

Table 4: Hierarchical regression analyses results.

Model		B	t	R^2	$\triangle R^2$
1	step1			0.38	0.38***
	Age	-0.12	-0.59		
	Sex	0.049	0.80		
	Grade	0.52	2.90**		
	Major	0.066	2.83**		
	General grit	0.76	10.35***		
	step2			0.51	0.13***
	Age	-0.247	-1.39		
	Sex	0.033	0.60		
	Grade	0.402	2.50*		
	Major	0.058	2.76**		
	General grit	0.455	5.94***		
	L2PoE	0.088	1.90*		
	L2CoI	0.199	4.93***		
	L2AtS	0.154	3.15**		
2	step1			0.31	0.29***
	Age	-0.431	-1.552		
	Sex	0.006	0.069		
	Grade	0.252	0.994		
	Major	0.075	2.262*		
	General grit	0.960	9.294***		
	step2			0.41	0.10***
	Age	-0.46	-1.731		
	Sex	-0.005	-0.064		
	Grade	0.14	0.609		
	Major	0.063	2.045		
	General grit	0.61	5.402***		
	L2PoE	0.11	1.648		
	L2CoI	0.070	1.176		
	L2AtS	0.27	3.725*		

The outcome variable of model 1 and 2 is L2 engagement, and L2 enjoyment, respectively. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

In the task of predicting L2 enjoyment (Table 5), employing L2GSC proved to be more advantageous than using TMGS. Specifically, in the performance of L2GSC, the Random Forest model achieved the highest accuracy at 72.92 %, precision at 75.50 %, recall at 72.92 %, and F1-score at 73.30 %. By contrast, the performance of TMGS, as highlighted by the LightGBM model, achieved the highest accuracy at 66.67 %, precision at 63.81 %, recall at 66.67 %, and F1-score at 65.20 %. It proved that the L2GSC evidently overweighed TMGS in predicting L2 enjoyment.

In terms of the analysis of predicting L2 engagement (Table 6), in the performance of L2GSC, the Random Forest model achieved accuracy at 70.83 %, precision at

Table 5: Model performance comparison in predicting L2 enjoyment using data from L2GSC and TMGS.

Predicting L2 Enjoyment: Model Performance Comparison (Weighted Average)								
	Performance of L2GSC				Performance of TMGS			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
RandomForest	72.92 %	75.50 %	72.92 %	73.30 %	60.42 %	58.89 %	60.42 %	59.58 %
XGBoost	68.75 %	73.49 %	68.75 %	70.09 %	60.42 %	60.84 %	60.42 %	60.50 %
LightGBM	68.75 %	74.14 %	68.75 %	69.40 %	66.67 %	63.81 %	66.67 %	65.20 %

Table 6: Model performance comparison in predicting L2 engagement using data from L2GSC and TMGS.

Predicting L2 Engagement: Model Performance Comparison (Weighted Average)								
	Performance of L2GSC				Performance of TMGS			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
RandomForest	70.83 %	75.20 %	70.83 %	72.84 %	77.08 %	73.51 %	77.08 %	75.25 %
XGBoost	72.92 %	75.49 %	72.92 %	74.08 %	77.08 %	73.51 %	77.08 %	75.25 %
LightGBM	70.83 %	67.46 %	70.83 %	69.11 %	77.08 %	75.25 %	77.08 %	76.16 %

75.20 %, recall at 70.83 %, and F1-score at 72.84 %. The TMGS performed best with the LightGBM model. It achieves an accuracy of 77.08 %, precision at 75.25 %, recall at 77.08 %, and F1-score at 76.16 %. This suggests that TMGS might be slightly more predictive of L2 engagement.

5 Discussion

The aim of this research was to present initial evidence supporting the reliability and validity underlying a newly developed L2 grit scale for measuring L2 grit within a collectivist culture. Overall, the findings indicate that the L2GSC scores are psychometrically robust, offering a credible measure of L2 grit within the context of a collectivist culture (i.e., China in this study).

By integrating EFA data from Phase 2 with CFA data in Phase 3, the research has affirmed the second-order three-factor model of grit, which further replicated the findings of Datu et al. (2017), who found similar factor structures in samples from collective societies. The 12-item L2GSC, encompassing three distinct factors (i.e., perseverance of effort, consistency of interest, and adaptability to situations),

was validated through evidence of strong psychometric properties, including internal and external validity and reliability. Furthermore, the scale's adequacy is highlighted by its composition of an appropriate number of items across three factors. In addition, this research not only distinguishes L2GSC from the general Triarchic Model of Grit by affirming and extending the conceptual coherence of Datu et al. (2017) but also sets it apart from other constructs such as L2-related grit scales described by Teimouri et al. (2020) and Alamer (2021), thereby contributing to the theoretical and empirical understanding of grit in SLA contexts from a collectivist/group-oriented perspective.

This work not only reaffirms the dual facets of PE and CI within the framework of L2 grit as outlined by Teimouri et al. (2020) but also corresponds with the Triarchic Model of Grit proposed by Datu et al. (2017). Aligning with the new dimension (i.e., adaptability to situations) proposed by Datu et al. (2017), an intriguing question arises regarding the role of adaptability within L2 learning. Pettigrew (1979) contends that culture is crafted by human action and, reciprocally, shapes human identity; that is, self-perception is significantly influenced by the cultural environment. Therefore, within the context of collectivist cultures in which individuals' learning habits are often underpinned by Confucian values, the construct of AS potentially serves as a critical component in the acquisition and application of an L2 for the following two reasons.

Specifically, the concept of Confucian dynamism refers to the extent to which a society emphasizes a future or long-term orientation or a past or short-term orientation and reflects the Confucian ethic of diligence (Hofstede and Bond 1988). Individuals from East Asian cultures are particularly noted for their strong success orientation, often placing a higher premium on hard work than typically observed in Western cultures (Cheng 1990). This study posits that effective L2 learning outcomes are a reflection of educational success. Such outcomes are frequently linked with attributes of perseverance and the ability to overcome challenges in the L2 learning journey, traits that are encapsulated by the term "grit". Consequently, the high value that culture places on achievement might inspire students to persist in their English language studies, despite encountering unforeseen obstacles. Second, Confucian culture emphasizes people's ability to cope with changes and challenges though individuals should respect tradition and maintain social order (Niu 2012). It posits that individuals ought to exhibit flexibility and open-mindedness, proactively engaging with and molding their surroundings, whilst also permitting environmental influences to inform their development (Niu 2012). This idea of adaptability and flexibility can be considered a response of Confucian culture to the constant change and unpredictability of real life. Therefore, in the L2 learning process, learners are encouraged to remain adaptable and responsive to each emerging opportunity.

In addition, this study examined the interplay between language learners' language growth mindsets, conscientiousness, engagement, enjoyment, and their association with L2 grit. The outcomes of the correlational analysis revealed that both PE and AS demonstrated positive correlations with language growth mindset while consistency is not. From the perspective of collectivism, people emphasize the importance of interpersonal relationships within groups and the extended family (Boucher 2010), which can be regarded as an "other-directed self" (Wen and Clément 2003, p. 19). The robust network of communal relations offers consistent support to individuals, who in turn are expected to demonstrate unwavering loyalty to their group (Tang 2012). In such contexts, personal identity is less about individual uniqueness and more about external, socially defined attributes like roles, status, and affiliations with family and other collectives (Boucher 2010). Therefore, L2 learners may exhibit a diligent pursuit of successful language acquisition outcomes, striving to overcome any challenges to meet the expectations of others, such as family members and teachers, which can be labeled as subjective norms (Ajzen and Fishbein 1975).

Furthermore, both PE and AS showed significant positive associations with conscientiousness, whereas CI had a minor positive association with the same trait. Teimouri et al. (2020) reported no correlation between conscientiousness and L2 grit, diverging from past research indicating a strong link between conscientiousness and general grit. It's important to consider future research to explore if L2 grit within a collective context is connected to conscientiousness in the same manner as general grit or conscientiousness has the potential to predict L2 grit within a collective culture, examining this in different environments.

The regression results showed that students with higher L2 grit were more attentively engaged in classroom activities and exhibited a more pronounced predilection for enjoying the language learning journey compared to their counterparts with lower grit. These outcomes align with the corpus of grit literature and L2 learning, which implies satisfactory L2 learning outcomes (Fathi and Hejazi 2024; Sadoughi and Hejazi 2023). This suggests that L2GSC is a reliable indicator of L2 learners' behavioral patterns, confirming its appropriateness for the L2 context.

Based on ML analysis, the results demonstrate that the L2GSC outperforms the TMGS in predicting L2 enjoyment. The superior predictive power of L2GSC suggests that it is a dependable measure for anticipating the behavioral trends of L2 learners within language learning contexts. An interesting finding was that TMGS outperformed L2GSC in predicting L2 engagement. While this aspect of the findings may seem ambiguous, three potential reasons could explain this. First, the participants in this study were primarily from China, where English is a mandatory subject in universities. This requirement aims to enhance their language proficiency, which in turn improves their employability post-graduation. As a result, their engagement in learning English may play an important role in fostering their future career success, which motivates them to dedicate time and effort to their studies. In this context,

students likely study diligently not out of interest in specific subjects, but rather with the general goal of achieving academic success across all courses. This suggests that their grit may be more generalized rather than specific to any one domain. However, future research should examine whether this finding holds true across different subjects and areas of study. In addition, we speculate that L2 enjoyment might be a more domain-specific variable than L2 engagement. This is supported by an analysis revealing that enjoyment is a more powerful predictor of achievement in L2 vocabulary than engagement (Oyama 2022), affirming the significant role of enjoyment in the learning process, making it a more suitable predictor of L2 grit. The varying degrees of domain specificity among different variables within the same domain is an intriguing question that deserves future research. This also demonstrates the robust predictive power of ML methods, which can help SLA researchers and educators better understand and recognize the underlying relationships between different language learning conditions. Third, this result was achieved through the innovative application of ML in scale validation, marking a significant methodological advancement. Comparing with using hierarchical regression in predicting engagement, ML provided distinct findings that deepen our understanding of the data. This approach offers a novel contribution, as prior research has typically relied on regression or correlation analyses for scale validation (e.g., Datu et al. 2017; Datu and Zhang 2021), rarely extending to the integration of ML techniques. By conducting a more comprehensive analysis with ML, we were able to reveal a new perspective that highlights the added value and necessity of employing diverse analytical methods to examine the same research question. This not only enhances the robustness of our findings but also underscores the potential of ML to refine and extend traditional validation processes.

In brief, the research has elucidated several key findings regarding the characteristics and outcomes associated with L2 grit. First, students exhibiting grit can demonstrate sustained persistence in their L2 learning endeavors, maintain a passion for engaging with the target language, and show an increased awareness of and responsiveness to evolving opportunities and changes within their L2 learning context. Second, gritty students can experience a heightened sense of engagement and derive more enjoyment from their L2 learning activities and may harbor the conviction that their language learning achievements can be significantly enhanced through increased effort. Therefore, it is clear that grit is not merely a peripheral trait but is central to fostering persistence, enthusiasm, and an adaptive response to the learning environment among students.

6 Limitations and future directions

This study encounters certain limitations. First, the self-reporting method may introduce biases, such as social desirability, and memory recall potentially affecting

data accuracy. Second, use of convenience sampling, while practical, may limit the representativeness of the dataset. Random sampling in future studies could address this issue. Also, to establish a more robust cultural foundation, incorporating students from Chinese Departments may yield stronger evidence of Chinese collectivism. Third, the study employed a cross-sectional design and did not include language achievement measures, such as English test scores. Future research could adopt longitudinal or mixed-method approaches to better assess the L2GSC's predictive capacity for language learning outcomes. Finally, this study represents a pioneering effort to create models using ML methods to evaluate scalability and predictability. The psychometric validation of the L2GSC was conducted within a single collectivist cultural context (China). Future research should gather more data to build more models and enhance predictive performance and test the L2GSC across diverse cultural settings to evaluate its cross-cultural applicability and replicability of findings.

7 Conclusions

The present study developed and validated L2GSC, a tool designed to assess grit in the context of collectivist culture specifically within SLA. The findings indicate that students exhibit higher grit levels in language learning scenarios than in general contexts when assessed with this scale. This suggests that a domain-specific measure of grit yields more definitive and significant outcomes. Building on prior studies, this work integrates the concept of L2 grit into the collectivist societal framework and investigates its impact on the learning behaviors of L2 learners. In short, the results suggested that grittier individuals tend to be more engaged and enjoyable in language learning compared to their less gritty peers.

This research contributes to theoretical understanding, practical application, and pedagogical implications. Theoretically, it enriches the construct of L2 grit with cultural nuances, potentially enhancing SLA processes. Practically, the L2GSC offers a valuable instrument for researchers and educators to measure individual variances in PE, CI, and AS toward long-term objectives in SLA. Should further replication studies consistently validate the factorial structure of the L2 grit scale across various collective contexts, this scale will become a valuable contribution to the current measures for grit specific to the language domain. Pedagogically, the belief in the power of perseverance of effort, consistency of interest, and adaptive approaches, a hallmark of gritty students, suggests a promising avenue for enhancing language learning outcomes. Therefore, these insights advocate for the incorporation of strategies to cultivate grit within language learning curricula, aiming to unlock the full potential of learners in mastering an L2.

Appendices

Appendix A: L2GSC

题项	非常不同意	不同意	不知道	同意	非常同意
	1	2	3	4	5
Persistence of effort:					
1.就学习英语而言,我是一个努力学习的人。 When it comes to English, I am a hard-working learner.					
2.只要我决定学好英语,无论遇到什么困难,我都会坚定地实现我的目标。 Now that I have decided to learn English, nothing can prevent me from reaching this goal.					
3.我是一个勤奋的英语学习者。 I am a diligent English language learner.					
4.无论遭遇任何挫折,我都坚决不放弃英语学习。 I will not allow anything to stop me from my progress in learning English.					
5.我会花很多时间和精力来提高我的英语薄弱点。 I put much time and effort into improving my English language weakness.					
Consistency of interest:					
6.我现在对学习英语的兴趣不如以前那么浓厚。[a] I am not as interested in learning English as I used to be.[a]					
7.我曾经对语言学习很感兴趣,但最近我的兴趣似乎有所减退。[a]					
8.我对英语学习的兴趣一直在变。[a]					
Adaptability to situations:					
9.我愿意探索英语学习的新方法。 I was obsessed with learning English in the past but have lost interest recently.[a]					
10.适时调整我的学习策略对于实现我的长期语言目标至关重要。 Adapting my study methods is vital to achieving my long-term English proficiency goals.					
11.我能够适应英语学习中的不同环境或挑战。 I am able to adjust to new English learning environments or challenges.					
12.我始终有动力去提高我的语言学习技能。 I am constantly driven to enhance my English learning competencies.					

[a]Item was reverse scored; Only Chinese items were used in this study.

Appendix B: Phase 1 interview questions (original Chinese)

1. 您能分享一下您从小到大在学习英语过程中遇到的最重大挫折是什么吗?这些挫折对您产生了哪些影响?
   Can you share what are the biggest setbacks you have encountered in learning English since childhood? What impact have these setbacks had on you?
2. 面对这些挫折,您采取了哪些具体的策略来克服困难并提升英语水平?这些策略对您来说效果如何?
   In the face of these setbacks, what specific strategies have you adopted to overcome difficulties and improve your English level? How effective are these strategies for you?
3. 在学习英语的过程中,您如何看待个人适应能力的作用?您觉得适应能力在应对学习挑战中扮演了什么样的角色?
   In the process of learning English, how do you view the role of personal adaptability? What role do you think adaptability plays in coping with learning challenges?
4. 您在学习英语的不同阶段(如童年、青年、成年)中,面对挫折时的心态和应对方式是否有变化?如果有,请详细描述这些变化。
   Has your mentality and coping style changed when facing setbacks in different stages of learning English (such as childhood, youth, and adulthood)? If so, please describe these changes in detail.
5. 您认为除了个人努力外,外部因素(如环境、教师、同伴等)在帮助您克服学习英语的挫折中起到了哪些作用?
   In addition to personal efforts, what role do you think external factors (such as environment, teachers, peers, etc.) play in helping you overcome setbacks in learning English?
6. 您有没有通过某些特定的习惯或方法来保持并持续提高您的英语水平?这些方法是否有效地帮助您适应新的学习环境或提高语言能力?
   Do you have any specific habits or methods to maintain and continuously improve your English level? Do these methods effectively help you adapt to the new learning environment or improve your language skills?

Note. Only Chinese items were used in this study.

Appendix C: Rating rubric in inter-rater agreement assessment in phase 2

Likert scale (1–4, where 1 = strongly disagree and 4 = strongly agree).

1. **Relevance**: Each item is assessed on how relevant it is to the construct being measured.

2. **Clarity**: Raters evaluate whether the item is clearly worded and easily understood. This might include checking for ambiguity, clarity of language, and the appropriateness of the terminology used.
3. **Comprehensiveness**: Raters check if the items collectively cover all aspects of the construct. Each item might be evaluated on whether it addresses a specific dimension of the construct.
4. **Redundancy***: Raters might assess whether any items are redundant or overly similar to other items, which could unnecessarily lengthen the scale without adding value.
5. **Difficulty***: In some cases, raters might evaluate the difficulty or complexity of the item, ensuring it's appropriate for the target population.

Note. * reverse scoring.

Research funding: This work was supported by Shanghai Chenguang Talent Program (20CG40), Shanghai Philosophy and Social Science Planning Youth Project (2020EYY004) and Universidade de Macau (MYRG2022-00055-FED).

References

Ajzen, Icek & Martin Fishbein. 1975. A bayesian analysis of attribution processes. *Psychological Bulletin* 82(2). 261–277.

Alamer, Abdullah. 2021. Grit and language learning: Construct validation of L2-grit scale and its relation to later vocabulary knowledge. *Educational Psychology* 41(5). 544–562.

Black, William C, Babin Babin J & Rolph E. Anderson. 2010. *Multivariate data analysis: A global perspective*. Pearson.

Boucher, Helen C. 2010. Understanding Western East Asian differences and similarities in self enhancement. *Social and Personality Psychology Compass* 4(5). 304–317.

Breiman, Leo. 2001. Random forests. *Machine Learning* 45. 5–32.

Chen, Tianqi & Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 785–794. San Francisco, CA, USA: ACM.

Cheng, Stephen K. K. 1990. Understanding the culture and behaviour of East Asians-a Confucian perspective. *Australian and New Zealand Journal of Psychiatry* 24(4). 510–515.

Cheng, Stephen K. K. 2021. The effect of language-specific grit and future self-guides on willingness to communicate in the second language. *English Teaching and Learning* 45(3). 1–19.

Cormier, Danielle L, John G. H. Dunn & Janice Causgrove Dunn. 2019. Examining the domain specificity of grit. *Personality and Individual Differences* 139. 349–354.

Datu, Jesus Alfonso D. & Jiahong Zhang. 2021. Validating the Chinese version of triarchic model of grit scale in technical–vocational college students. *Journal of Psychoeducational Assessment* 39(3). 381–387.

Datu, Jesus Alfonso D, Jana Patricia M. Valdez & Ronnel B. King. 2016. Perseverance counts but consistency does not! Validating the short grit scale in a collectivist setting. *Current Psychology* 35. 121–130.

Datu, Jesus Alfonso D., Mantak Yuen & Gaowei Chen. 2017. Development and validation of the triarchic model of grit scale (TMGS): Evidence from Filipino undergraduate students. *Personality and Individual Differences* 114. 198–205.

Deci, Edward L. & Richard M. Ryan. 1985. The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality* 19(2). 109–134.

Dewaele, Jean-Marc & Chengchen Li. 2021. Teacher enthusiasm and students' social-behavioral learning engagement: The mediating role of student enjoyment and boredom in Chinese EFL classes. *Language Teaching Research* 25(6). 922–945.

Duckworth, Angela Lee & Patrick D. Quinn. 2009. Development and validation of the short grit scale (GRIT–S). *Journal of Personality Assessment* 91(2). 166–174.

Duckworth, Angela Lee, Christopher Peterson, Michael D. Matthews & Daniel R. Kelly. 2007. Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology* 92(6). 1087.

Eren, Altay & Anıl Rakıcıoğlu-Söylemez. 2023. Language mindsets, perceived instrumentality, engagement and graded performance in English as a foreign language students. *Language Teaching Research* 27(3). 544–574.

Eskreis-Winkler, Lauren, Elizabeth P. Shulman, Scott A. Beal & Angela Lee Duckworth. 2014. The grit effect: Predicting retention in the military, the workplace, school, and marriage. *Frontiers in Psychology* 5. 36.

Fathi, Jalil & S. Yasaman Hejazi. 2024. Ideal L2 self and foreign language achievement: The mediating roles of L2 grit and foreign language enjoyment. *Current Psychology*. 1–15. https://doi.org/10.1007/s12144-023-05187-8.

Fathi, Jalil, Mirosław Pawlak, Sam Saeedian & Ata Ghaderi. 2024. Exploring factors affecting foreign language achievement: The role of growth mindset, self-efficacy, and L2 grit. *Language Teaching Research*. https://doi.org/10.1177/13621688241227603.

Field, Andy 2024. *Discovering statistics using IBM SPSS statistics*. Sage Publications Limited.

Fredrickson, Barbara Lee 2003. The value of positive emotions: The emerging science of positive psychology is coming to understand why it's good to feel good. *American Scientist* 91(4). 330–335.

Hofstede, Geert & Michael Harris Bond. 1988. The Confucius connection: From cultural roots to economic growth. *Organizational Dynamics* 16(4). 5–21.

Hutcheson, Graeme D. & Sofroniou, Nick. 1999. The multivariate social scientist: Introductory statistics using generalized linear models. *The Multivariate Social Scientist*. 1–288.

Jiang, Yifan 2020. *A study on the multi-level impact of personality traits and organizational innovation situations on employee career growth*. Doctoral dissertation. China: University of Science and Technology of China.[

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye & Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* 30. https://proceedings.neurips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Khajavy, Gholam Hassan & Elham Aghaee. 2024. The contribution of grit, emotions and personal bests to foreign language learning. *Journal of Multilingual and Multicultural Development*. 1–15. https://doi.org/10.1080/01434632.2022.2047192.

Khajavy, Gholam Hassan, Peter David MacIntyre & Jamal Hariri. 2021. A closer look at grit and language mindset as predictors of foreign language achievement. *Studies in Second Language Acquisition* 43(2). 379–402.

Khan, Bisma Mazhar & Ali Mahmood Khan. 2017. Grit, happiness and life satisfaction among professionals: A correlational study. *Journal of Psychology and Cognition* 2(2).

Lan, Guoxiang, Larisa Nikitina & Wei Sheng Woo. 2021. Ideal L2 self and willingness to communicate: A moderated mediation model of shyness and grit. *System* 99. 102503.

Li, Chengchen & Yuan Yang. 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*(0). https://doi.org/10.1515/iral-2022-0196.

Li, Chengchen, Guiying Jiang & Jean-Marc Dewaele. 2018. Understanding Chinese high school students' foreign language enjoyment: Validation of the Chinese version of the foreign language enjoyment scale. *System* 76. 183–196.

Li, Chengchen, Li Wei & Jinting Cai. 2024. Life satisfaction and L2 engagement in adolescents. *ELT Journal*. ccad054.

Lin, Zhiqing. 2021. A Methodological Review of Machine Learning in Applied Linguistics. *English Language Teaching* 14(1). 74-85.

Littlewood, W. 1999. Defining and developing autonomy in East Asian contexts. *Applied Linguistics* 20(1). 71–94.

Liu, Eerdemutu, Junju Wang & Sachurina Bai. 2022. Self-guides, enjoyment, gender, and achievement: A survey of Chinese EFL high school students. *Journal of Multilingual and Multicultural Development*. 1–18. https://doi.org/10.1080/01434632.2022.2153854.

Liu, Honggang, Xiaoxue Li & Yan Yan. 2023a. Demystifying the predictive role of students' perceived foreign language teacher support in foreign language anxiety: The mediating role of L2 grit. *Journal of Multilingual and Multicultural Development*. https://doi.org/10.1080/01434632.2023.2223171.

Liu, Honggang, Yuchen Zhong, Huaruo Chen & Ying Wang. 2023b. The mediating roles of resilience and motivation in the relationship between students' English learning burnout on engagement: A conservation-of-resources perspective. *International Review of Applied Linguistics in Language Teaching*. https://doi.org/10.1515/iral-2023-0089.

Niemiec, Christopher P. & Richard M. Ryan. 2009. Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education* 7(2). 133–144.

Niu, Weihua 2012. Confucian ideology and creativity. *Journal of Creative Behavior* 46(4). 274–284.

Oyama, Ren 2022. Exploring the relationships between enjoyment, self-efficacy, engagement, and vocabulary learning in Japanese learners of English. *Journal of Asia TEFL* 19(4). 1163–1180.

Padilla, Amado M, Xinjie Chen & J. Lake. 2020. Positive psychology and learning a second or third language [Editorial]. *Frontiers in Psychology* 11. https://doi.org/10.3389/fpsyg.2020.599326.

Pettigrew, Andrew M. 1979. On studying organizational cultures. *Administrative Science Quarterly* 24(4). 570–581.

Rezaii, Neguine, Elaine Walker & Phillip Wolff. 2019. A Machine Learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophrenia* 5(1). 9.

Ryan, Richard M. & Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55(1). 68–78.

Sadoughi, Mehdi & S. Yasaman Hejazi. 2023. Teacher support, growth language mindset, and academic engagement: The mediating role of L2 grit. *Studies in Educational Evaluation* 77. 101251.

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

Shatte, Adrian B., Delyse M. Hutchinson & Samantha J. Teague. 2019. Machine Learning in mental health: A scoping review of methods and applications. *Psychological Medicine* 49(9). 1426–1448.

Sudina, Ekaterina & Luke Plonsky. 2021a. Academic perseverance in foreign language learning: An investigation of language-specific grit and its conceptual correlates. *The Modern Language Journal* 105(4). 829–857.
Sudina, Ekaterina & Luke Plonsky. 2021b. Language learning grit, achievement, and anxiety among L2 and L3 learners in Russia. *ITL - International Journal of Applied Linguistics* 172(2). 161–198.
Steger, Michael F., Patricia Frazier, Shigehiro Oishi & Matthew Kaler. 2006. The meaning in life questionnaire: assessing the presence of and search for meaning in life. *Journal of Counseling Psychology* 53(1). 80.
Sudina, Ekaterina, Jason Brown, Brien Datzman, Yukiko Oki, Katherine Song, Robert Cavanaugh, et al. 2021. Language-specific grit: Exploring psychometric properties, predictive validity, and differences across contexts. *Innovation in Language Learning and Teaching* 15(4). 334–351.
Suh, Eunkook M. 2007. Downsides of an overly context sensitive self: Implications from the culture and subjective well-being research. *Journal of Personality* 75(6). 1321–1343.
Tang, Mingyan 2012. Core spirit and function in Confucianism from the perspective of soft power of Chinese culture. *Cross-Cultural Communication* 8(6). 65.
Teimouri, Yasser, Luke Plonsky & Farhad Tabandeh. 2020. L2 grit: Passion and perseverance for second-language learning. *Language Teaching Research* 26(5). 893–918.
Von Culin, Katherine R., Eli Tsukayama & Angela Lee Duckworth. 2014. Unpacking grit: Motivational correlates of perseverance and passion for long-term goals. *The Journal of Positive Psychology* 9(4). 306–312.
Wen, Wei-Ping & Richard Clément. 2003. A Chinese conceptualisation of willingness to communicate in ESL. *Language Culture and Curriculum* 16(1). 18–38.
Wu, Chenggang, Xiaoyong Tian & Hui Jin. 2024. Grit and Second Language learning engagement: The mediating role of affect balance. *Behavioral Sciences* 14(3). 184. https://www.mdpi.com/2076-328X/14/3/184.
Yao, Yuan, Weijun Wang & Xiaopei Yang. 2021. Perceptions of the inclusion of automatic writing evaluation in peer assessment on EFL writers' language mindsets and motivation: A short-term longitudinal study. *Assessing Writing* 50. 100568.
Yarkoni, Tal & Jacob Westfall. 2017. Choosing prediction over explanation in psychology: Lessons from Machine Learning. *Perspectives on Psychological Science* 12(6). 1100–1122.
Zhao, Xian & Danping Wang. 2023. Grit in second language acquisition: A systematic review from 2017 to 2022. *Frontiers in Psychology* 14. https://doi.org/10.3389/fpsyg.2023.1238788.
Zhao, Yanli 2023. On the relationship between second language learners' grit, hope, and foreign language enjoyment. *Heliyon* 9(3). e13887.

Bionotes

Kenan Gao
Faculty of Education, University of Macau, Macao, China
yc37104@connect.um.edu.mo
https://orcid.org/0009-0004-7476-9514

Kenan Gao is a PhD student in Faculty of Education and Center for Cognitive and Brain Sciences, University of Macau. Her research centers on exploring innovative methods for vocabulary learning and integrating technology to improve the learning experience for English as a Second/Foreign Language (ESL/EFL)

students. Key areas of interest include technology-assisted language acquisition, learner-focused EFL learning approaches, and positive psychology.

Juan Zhang
Faculty of Education, University of Macau, Macao, China
juanzhang@um.edu.mo
https://orcid.org/0000-0002-7052-1093

Juan Zhang, PhD, is an associate professor and director of Centre for Early Childhood Education and Child Development at Faculty of Education, University of Macau. She uses behavioral and neuroscience methodologies (including EEG/ERP, fNIRs and fMRI) to investigate the brain mechanism underlying children's language and cognitive processes as well as social-emotional development. Key areas of interest include Language and Literacy Acquisition, Dyslexia, Bilingualism; Social-Emotional Development, Attachment, Parenting; Autism; STEM Education.

Yihui Wang
Faculty of Education, University of Macau, Macao, China
yc17105@connect.um.edu.mo
https://orcid.org/0000-0002-3221-3313

Yihui Wang is a PhD candidate in Faculty of Education and Center for Cognitive and Brain Sciences, University of Macau. Her research centers on investigating the dynamics of parent-child interactions and the neural mechanisms underlying these processes. Key areas of interest include fathers' and mothers' emotional socialization practices and their role in shaping children's understanding of social cues, emotional expressions, empathy, and theory of mind abilities at both the behavioral and neural levels. Her work leverages fNIRS-based hyperscanning to examine neural synchronization during parent-child interactions.

Chenggang Wu
School of Education, Shanghai International Studies University, Shanghai, China
chenggangwu@shisu.edu.cn
https://orcid.org/0000-0003-3837-3841

Chenggang Wu, PhD, is a lecturer in School of Education and Key Laboratory of Multilingual Education with AI, Shanghai International Studies University. His research areas focus on psycholinguistics, neurolinguistics, educational psychology and language learning.

Wei He
Department of Engineering Science, Macau University of Science and Technology, Macao, China
vicchoo@outlook.com
https://orcid.org/0000-0001-7786-2715

Wei He is a PhD student at the Department of Engineering Science, Faculty of Innovation Engineering, Macau University of Science and Technology. His research interests are artificial intelligence, deep learning, and time series analysis.