International Journal of Assessment Tools in Education
2026, Vol. 13, No. 2, 359–378
https://doi.org/10.21449/ijate.1767979
journal homepage: https://dergipark.org.tr/en/pub/ijate *Research Article*

Domain-specific approach to grit: A scale adaptation in the context of Turkish as a foreign language

Enes Çinpolat[1*], **Muhammed Eren Uygur**[1]

[1]Ordu University, Faculty of Education, Department of Turkish and Social Sciences, Ordu, Türkiye

ARTICLE HISTORY

Received: Aug. 26, 2025

Accepted: Nov. 09, 2025

KEYWORDS

L2 grit,
Turkish as a foreign language,
Scale adaptation,
Criterion validity,
Measurement invariance.

Abstract: This study aims to adapt and validate the L2 Grit Scale for the Turkish as a foreign language context. Data were collected from four distinct sample groups (N = 658) of university preparatory program students using a convenience sampling method. The adaptation process involved a back-translation procedure with two translators and an English Language Teaching expert, followed by a pilot test and linguistic equivalence confirmation. Validity analyses confirmed the two-factor structure through Confirmatory Factor Analysis (CFA) (CFI = .96, RMSEA = .07, SRMR = .03) and Exploratory Structural Equation Modelling (ESEM) (CFI = .97, RMSEA = .05, SRMR = .02). Convergent validity was supported by satisfactory average variance extracted (AVE > .53) and composite reliability (CR > .83) values. Furthermore, L2 Grit emerged as a stronger predictor of willingness to communicate in Turkish (WTC) than general Grit, emphasizing its distinct role in language learning motivation. The scale's validity was further confirmed by significant correlations with foreign language enjoyment (FLE), willingness to communicate (WTC), foreign language peace of mind (FLPOM), and foreign language classroom anxiety (FLCA). Additionally, L2 Grit demonstrated a significant relationship with students' exam scores, confirming its relevance in assessing perseverance and success in language learning. Measurement invariance was established across proficiency levels and genders. Reliability analyses indicated strong internal consistency (Cronbach's α = .76, McDonald's ω = .77). The results demonstrate that the Turkish version of the L2 Grit Scale is a valid and reliable instrument for learners of Turkish as a foreign language.

INTRODUCTION

The impact of knowledge, intelligence, interest, motivation, and anxiety on language learning processes is well-documented. Beyond these factors, attributes like passion, perseverance, resilience, and persistence are sometimes even more influential than IQ, talent, or prior knowledge, especially in long-term pursuits such as language learning (Alamer, 2021; Duckworth *et al.*, 2007; Mikami, 2024; Teimouri *et al.*, 2020). Duckworth (2017) coined "grit" to describe a construct encompassing these qualities, defining it as "perseverance and passion for long-term goals" (p. 1087). This construct refers to an individual's capacity to maintain passion and perseverance toward long-term goals and consists of two core dimensions: *Perseverance of Effort* (PoE) and *Consistency of Interest* (CoI) (Fathi *et al.*, 2024, p. 6; Liu, 2022, p. 2). PoE refers to the ability to maintain effort for a long time in the face of difficulty

*Corresponding author at: Enes ÇİNPOLAT ✉ enescinpolat@gmail.com Ordu University, Faculty of Education, Department of Turkish and Social Sciences, Ordu, Türkiye

e-ISSN: 2148-7456

or failure, whereas CoI defines the capacity to sustain interest in a particular goal over the long term despite challenges or setbacks (Pawlak *et al*., 2022, pp. 2094-2095). Initially introduced in positive psychology, the concept has become central to research on language learning, as it is seen as a critical non-cognitive factor that can predict success beyond innate talent (Wang, 2021, p. 1).

The importance of grit in language learning is increasingly recognized, with research in this area having "entered a fast growth phase since 2020" (Zhao & Wang, 2023, p. 1). A systematic review by Zhao and Wang (2023), which covered 32 empirical articles, and other meta-analytic syntheses (e.g., Credé *et al*., 2017), confirm that grit is a critical factor for long-term success and is often a stronger predictor of academic outcomes than talent (Li & Yuan, 2024, p. 5; Wang, 2021, p. 2). However, these comprehensive reviews also highlight significant gaps in the literature that this study aims to address. First, existing research has largely focused on English language learning, and the "lack of grit research in other languages" hinders a comprehensive understanding of the concept in different linguistic contexts (Zhao & Wang, 2023, p. 7). Second, there is an ongoing debate about the conceptualization and factor structure of grit, and there is a clear call that validate and extend both popular scales across different languages and contexts are further warranted to clarify its function in Second Language (L2) learning (Zhao & Wang, 2023, p. 6). Therefore, this study is important as it directly responds to these calls by adapting and validating a domain-specific L2 grit scale for the under-researched context of Turkish as a foreign language. By doing so, it will provide a psychometrically sound tool for future research and contribute to a more nuanced, cross-linguistic understanding of how passion and perseverance contribute to success in language learning.

As studies on grit in the L2 domain have expanded, researchers have distinguished between domain-general and domain-specific grit. Domain-general grit references a broader application of grit across various areas, while domain-specific grit, as applied in this study, pertains specifically to second language learning. Grit comprises two core dimensions: perseverance of effort and consistency of interest. Grit manifests through commitment and persistence in pursuing long-term goals (Morell *et al.*, 2021). Despite technological advances making language learning seemingly effortless, it remains a process that demands sustained effort and resilience (Bi *et al.*, 2024). For language learners who struggle to persist, factors such as "lack of motivation, remaining within one's comfort zone, extended breaks, insufficient review, and prioritizing other needs" (Gurbuz, 2022, p. 323), along with issues related to "classroom environment, instructor limitations, and insufficient instructional activities" (Evans & Tragant, 2020; Javadi Safa *et al.*, 2022), often contribute to the likelihood of abandoning language learning. Therefore, grit appears essential for long-term success in learning a foreign language (Khajavy *et al.*, 2020; Teimouri *et al.*, 2020; Wang, 2021).

Field research reveals that studies on grit in language learning predominantly focus on English and have increased since the 2020s, consistently highlighting grit as a critical factor for success in language acquisition over extended periods (Credé *et al.*, 2017; Demir, 2024; Sun & Xu, 2024). Domain-specific studies examining the relationship between grit and academic success indicate that grit is a significant predictor of L2 success (Sun & Xu, 2024) and may correlate with factors like motivation, willingness to communicate, engagement, and enjoyment (Alrabai, 2024; Ismail *et al.*, 2023; Lee, 2022; Yang *et al.*, 2024). Furthermore, it demonstrates meaningful mediation effects with variables like academic burnout and anxiety (Xu *et al.*, 2022). Researchers have also explored the relationship between domain-general and domain-specific grit. Some studies have found moderate to low correlations within the L2 context (Alrabai, 2024; Botes *et al.*, 2024), while others have noted an absence of linear association between domain-general and domain-specific grit (Li & Yang, 2023).

Studies examining the impact of grit in second language learning beyond English remain limited (Demir, 2024; Sun & Xu, 2024; Zhao & Wang, 2023). Also, in other languages there are very limited but significant studies on both domain-general and domain-specific exist (Demir, 2024; Rost *et al.*, 2024). Existing research, however, shows similar results regarding its predictive power in English learning contexts. Currently, there are no studies on grit specific to learning Turkish, and research within Türkiye has focused primarily on adapting grit scales within the EFL context or general grit (Akın & Arıcı, 2015; Sarıçam *et al.*, 2016; Taşpınar & Külekçi, 2018; Uştuk & Erarslan, 2023). For example, Akın & Arıcı (2015) adapted the Short Grit Scale, while Sarıçam *et al.* (2016) conducted a similar adaptation with consistent findings. Taşpınar & Külekçi (2018) produced a corpus-based study highlighting the importance of grit in the EFL context, and Uştuk & Erarslan (2023) adapted Teimouri *et al.*'s (2020) L2 Grit Scale within the same EFL context in Turkish.

Research on language learning frequently explores emotions, affective responses, motivation, and factors influencing motivation. However, while related to these factors, the concept of grit is distinct and functions independently within its dynamics. Studies on affective factors in Turkish contexts reveal research focusing on anxiety, enjoyment, and emotional intelligence in language learning (Ateş, 2019; Ozer & Altay, 2021; Şakrak, 2009; Yeşilçınar & Erdemir, 2023), suggesting an interrelationship between affective dimensions and grit. In line with the best practices for cross-cultural research, a critical decision was made regarding whether to develop a new scale or adapt an existing one (Hambleton & Patsula, 1999). Instead of developing a new measurement tool, adapting an existing one was deemed more practical and theoretically sound. This approach enables comparability with a growing body of international studies on L2 grit, ensures theoretical consistency with the established construct, and provides a tested factor structure that can be transferred into the Turkish as a foreign language context. This decision also addresses the fundamental question of whether the construct of "grit" is equivalent and meaningful across both the source and target cultures, a prerequisite for successful adaptation (Hambleton *et al.*, 2005). Accordingly, the present study aims to adapt L2 grit specifically for Turkish as a foreign language, examining the distinctions between domain-general and domain-specific grit while analyzing relationships with Foreign Language Enjoyment, Foreign Language Classroom Anxiety, and Willingness to Communicate. Just as enjoyment, anxiety, and willingness to communicate impact long-term language acquisition, grit has also been studied these variables (Demir, 2024; Zhao & Wang, 2023). Examining the correlation between these constructs and distinguishing domain-general from domain-specific grit offers valuable insights into the role of grit in language learning, particularly within the context of teaching Turkish as a foreign language.

While research on L2 grit has expanded, it remains predominantly focused on the context of English language learning. Studies examining grit in the acquisition of other foreign languages, particularly Turkish, are notably absent from the literature. This study addresses this critical gap by providing the first psychometrically validated instrument for this context. Specifically, this study adopts the two-factor conceptualization of grit (comprising Perseverance of Effort and Consistency of Interest) originally proposed by Duckworth *et al.* (2007). The operationalization of this construct for the L2 context, and the specific instrument adapted herein, is based on the L2-Grit Scale developed by Teimouri *et al.* (2020). By validating this instrument, the research offers a robust tool for exploring perseverance and passion in Turkish language acquisition. Therefore, this study seeks to answer the following research questions:

1. Is the Turkish version of the L2 Grit Scale linguistically equivalent to the original English version?

2. Does the two-factor structure (Perseverance of Effort and Consistency of Interest) of the L2 Grit Scale demonstrate good construct validity in the Turkish as a foreign language context?

3. Does the scale show adequate criterion and discriminant validity through its relationships with general grit, exam scores, and other affective variables?

4. Is the scale a reliable instrument for measuring foreign language specific grit among learners of Turkish?

5. Does the scale demonstrate measurement invariance across different language proficiency levels (B1, B2, C1) and genders?

METHOD

Design

This study utilized a survey research design to adapt and psychometrically validate the Turkish version of the L2 Grit Scale (Fraenkel *et al*., 2022). The process involved multiple quantitative steps, including correlational analyses and factor analyses, to establish the reliability and validity of the Turkish L2 Grit Scale.

Participants

The data were collected from four distinct student groups to systematically address different aspects of the scale's validation using a convenience sampling method (Fraenkel *et al.*, 2022). Each sample served a specific purpose: Sample 1 was used for linguistic validity; Sample 2 for initial construct validity (Confirmatory Factor Analysis - CFA); and Samples 3 and 4 for criterion, discriminant, and cross-validation analyses (Exploratory Structural Equation Modelling - ESEM). Participants were selected from the preparatory departments of four Turkish state universities. These preparatory programs are intensive, year-long courses designed to bring international students' Turkish proficiency to a level sufficient for academic study. The primary selection criteria for all participants were their enrollment in these programs, which utilize a standardized teaching framework (Common European Framework of Reference for Languages [CEFR]) and common course materials (Yeni İstanbul Turkish for International Students), ensuring a comparable baseline of language instruction across samples.

Sample 1: There are 18 university students in the language validity study. Ten identified themselves as male (55.5%) and eight as female (44.4%). The self-report determined that the English language level of 11 students was C1, 2 were B2, and 5 were B1. 16 people reported their Turkish language level as C1, 1 as B2, and 1 as B1.

Sample 2: 234 university students were included in the construct validity study group. 136 (58.1%) of the students reported their gender as male, and 98 (41.8%) as female. Based on their last successfully completed course and confirmed by self-report, 118 (50.4%) of the participants had C1, 71 (30.3%) had B2, and 45 (19.2%) had B1-level Turkish proficiency.

Sample 3: This sample consists of 186 university students. 115 (61.8%) of the students stated their gender as male and 70 (37.6%) as female. One person did not specify their gender. Based on their last successfully completed course and confirmed by self-report, 77 (41.3%) of the participants had C1, 49 (26.3%) had B2, and 60 (32.2%) had B1-level Turkish proficiency.

Sample 4: This sample consists of 220 university students learning Turkish as a foreign language. Among the participants, 89 (40.5%) identified as male, 128 (58.2%) as female, and 3 (1.4%) preferred not to disclose their gender. Based on their last successfully completed course and confirmed via self-report, 137 (62.3%) were at the B1 level, 58 (26.4%) at B2, and 25 (11.4%) at C1.

It is important to compare the sample of this study with that of the original scale development. The L2-Grit Scale was originally validated with 191 Farsi-speaking EFL learners in Iran (Teimouri *et al*., 2020). The sample in the current study is similar in that it also consists of university-level language learners engaged in intensive study. However, it differs in two key aspects: the target language is Turkish (not English), and the participants come from diverse linguistic and cultural backgrounds rather than a single L1 group. This diversity could make the

current validation particularly valuable for assessing the scale's robustness in a multicultural foreign language learning context.

Measures

L2-Grit Scale: The L2-Grit Scale was developed by Teimouri *et al.* (2020) to measure the language grit levels of English as a foreign language (EFL) learners (sample: 191 EFL students). It is a five-point Likert with nine items. The scale has a two-factor structure: perseverance of effort and consistency of interest. Perseverance of effort is loaded with five items, while consistency of interest is loaded with four items. Possible scores range from nine to 45. The scale has items such as "I am a diligent English language learner' and 'My interests in learning English change yearly." Various validity studies have also been conducted in teaching English as a foreign language (Sudina *et al.*, 2021; Uştuk & Erarslan, 2023; Wei *et al.*, 2020). The adapted Turkish version used in this study consists of nine items rated on a five-point Likert scale with the following anchors: 1 = *Not like me at all*, 2 = *Not much like me*, 3 = *Some how like me*, 4 = *Mostly like me*, and 5 = *Very much like me*. The scale has a two-factor structure: Perseverance of Effort (5 items; e.g., "I am a diligent Turkish language learner") and Consistency of Interest (4 items; e.g., "My interests in learning Turkish change yearly/monthly"). Scoring: The four items in the Consistency of Interest subscale (items 2, 4, 7, and 8) are reverse-scored (i.e., a response of 1 becomes 5, 2 becomes 4, etc.). A total L2-Grit score is then calculated by taking the average of all nine item scores. Higher scores indicate a higher level of L2-specific grit.

Grit-Short Form (Grit-S): The Grit-S, developed by Duckworth and Quinn (2009), was used to measure domain-general grit. The Turkish adaptation was conducted by Sarıçam *et al*. (2016). The scale consists of eight items on a five-point Likert scale (1 = *Not at all like me* to 5 = *Very much like me*). It comprises two factors: Perseverance of Effort and Consistency of Interest. The four Consistency of Interest items are reverse-scored. The total score is calculated by averaging all eight items. The Turkish version has demonstrated good reliability (Cronbach's α = .83).

Foreign Language Enjoyment Scale-Short Form (S-FLES): The FLE was first introduced as a 21-item scale by Dewaele and MacIntyre (2016) and later shortened to a 9-item version by Botes *et al*. (2021). The Turkish version of this short form, adapted and validated for Turkish learners of *English* by Yeşilçınar *et al*. (2025), was used in this study. For the context of the present study, all items, which were prefaced with "In my English class..." in the Turkish adaptation, were modified to refer to the Turkish class. A CFA on the current sample data (Sample 4, n = 220) supported the three-factor structure with excellent model fit: χ^2 / df = 1.29, CFI = .99, TLI = .98, RMSEA = .03, SRMR = .03. All items are scored on a five-point Likert scale (1 = *Strongly Disagree* to 5 = *Strongly Agree*) and averaged to create a total score. The scale showed high internal consistency in the original Turkish adaptation (Cronbach's α = .85) and in the current sample (Cronbach's α = .85).

Foreign Language Classroom Anxiety Scale-Short Form (S-FLCAS): The Turkish adaptation (Yeşilçınar *et al*., 2025) of the S-FLCAS (originally developed by Horwitz *et al*., 1986; shortened by Botes *et al*., 2022) was used. This 8-item scale was validated with Turkish learners of English. For the present study, items were adapted to refer to the Turkish class. A CFA on the current sample data (Sample 4, n = 220), with a modification between the error covariances of items 4 and 5, confirmed the single-factor structure with good model fit: χ^2 / df = 1.99, CFI = .96, TLI = .95, RMSEA = .06. Two items (Item 4 and Item 5) are reverse-scored. A total score is calculated by averaging all items. The scale showed high reliability in the prior Turkish adaptation (Cronbach's α = .93) and acceptable reliability in the current sample (Cronbach's α = .80).

Foreign Language Peace of Mind Scale (FLPOM): The Turkish adaptation (Yeşilçınar *et al*., 2025) of the FLPOM scale, originally developed by Zhou *et al*. (2023), was used. The scale was originally validated for Turkish learners of English. For this study, items were adapted to refer to learning Turkish. A CFA on the current sample data (Sample 4, n = 220) indicated an acceptable model fit for the single-factor structure: χ^2 / df = 2.91, CFI = .93, TLI = .90, RMSEA = .09, SRMR = .05. All items are scored directly by averaging the responses. The scale demonstrated excellent reliability in the original adaptation (Cronbach's α = .91) and in the current sample (Cronbach's α = .83).

Willingness to Communicate Scale (WTC): The Turkish version of the WTC scale, specifically adapted for the context of teaching Turkish as a foreign language (Çinpolat *et al*., 2024) from the original by Lee and Drajati (2020) was used. The adapted scale consists of 12 items assessing WTC across three contexts: In-Class, Out-of-Class, and Informal Digital Environment. Responses are given on a five-point Likert scale with specific anchors: "1 (*Definitely not willing*) to 5 (*Definitely willing*)". All items are scored directly, and a total score is calculated by averaging the items. The Turkish adaptation has demonstrated a valid three-factor structure and high internal consistency in its validation study (Cronbach's α = .87).

Exam Score: Participants' exam scores were obtained from the standardized course-level completion exams used in the Yeni İstanbul Turkish for International Students teaching set, which is implemented in the preparatory programs of the participating universities. These exams are designed in alignment with the CEFR and serve as the primary assessment for determining students' readiness to progress to the next course level. The exam's total score is 100. The use of a standardized exam across institutions ensures consistency in assessing participants' Turkish language proficiency.

Procedure

Scale adaptation process

The adaptation process meticulously followed the guidelines for cross-cultural scale adaptation (Hambleton & Patsula, 1999; Hambleton *et al.*, 2005). The procedure was conducted in the following stages:

1. Permission: Formal permission to translate and adapt the L2 Grit scale for research purposes was obtained from the corresponding author of the original study.

2. Translation and Back-Translation: A back-translation procedure was employed to ensure linguistic accuracy (Brislin, 1970). First, the scale was translated from English into Turkish by a professional translation graduate. Subsequently, this Turkish version was translated back into English by a separate translator with a degree in English Language Teaching (ELT) who was blind to the original English version.

3. Expert Panel Review: A panel consisting of an expert with a PhD in ELT, an expert with a PhD in Teaching Turkish as a Foreign Language, and the researchers themselves reviewed the original scale, the Turkish translation, and the back-translated version. The panel, whose members are familiar with the principles of scale adaptation, assessed four key types of equivalence (Kennedy, 2005). First, semantic equivalence was reviewed to ensure that words and sentence structures held the same meaning across languages. Second, the panel checked for idiomatic equivalence, addressing any colloquial expressions that might not translate directly. A significant focus was placed on experiential equivalence, which confirms that the situations described in the items are relevant and meaningful to the target culture. In this regard, a key decision was made: the item "My interests in learning Turkish change from year to year" was adapted to "change yearly/monthly". This change was deemed necessary because the target population consists of preparatory program students who typically have been learning Turkish for less than a year, making the original "yearly" timeframe less relevant to their experience.

Lastly, the panel verified conceptual equivalence, ensuring that the underlying psychological construct of "grit" is understood in the same way across cultures.

4. Cognitive Interviewing (Pilot Testing): The refined Turkish version was administered to five target language learners. They were asked to think aloud as they answered each item and provide feedback on the clarity, comprehensibility, and cultural appropriateness of the items. Their feedback confirmed that the items were well-understood and no major revisions were necessary.

5. Linguistic Equivalence Check: A bilingual group design was used to statistically verify language equivalence. A group of 18 participants proficient in both English and Turkish completed both the English version (adapted to refer to "Turkish") and the final Turkish version with a two-week interval between administrations. This step provided the data for the Wilcoxon signed-rank test reported in the Results section.

6. Proceeding to Main Psychometric Validation: The successful completion of the preliminary stages (namely, positive expert panel review, confirmation of comprehensibility in the pilot test, and the establishment of statistical linguistic equivalence) served as the criteria for proceeding to the main study. Having thus established the quality of the adapted instrument, data were collected from the main study groups (Samples 2, 3, and 4) to conduct comprehensive psychometric analyses. These analyses, which include construct validity (CFA, ESEM), criterion and discriminant validity, measurement invariance, and internal consistency reliability (α and ω), form the core of the study and are detailed in the Data Analysis and Results sections.

Main data collection

This research was conducted in accordance with ethical principles, with ethical approval obtained from the Education Research Ethics Committee of Ordu University on 18/07/2024 under approval number 2024-127. After receiving approval and successfully completing the translation process, data for the main study were collected online via Google Forms from participants in Samples 2, 3, and 4. The data collection was supervised by researchers to ensure data quality and to address any participant inquiries.

Data Analysis

All statistical analyses were conducted using JASP (Version 0.19.3.0) and Jamovi (Version 2.6.25). Before the main analyses, the dataset was screened for accuracy and missing values. The statistical assumptions for each procedure were explicitly examined to ensure the appropriateness of the selected analyses. For linguistic equivalence (Sample 1), the data were first assessed for normality using the Shapiro-Wilk test. As the assumption of normality was violated ($p < .05$), the non-parametric Wilcoxon signed-rank test was chosen to compare the scores from the English and Turkish forms of the scale.

For construct validity analyses (CFA and ESEM), preliminary data screening was conducted. Univariate normality was deemed acceptable as skewness and kurtosis values for all items were within the ±1.5 range, and the absence of significant univariate outliers was confirmed by inspecting z-scores, all of which were within the ±3 range. Despite these acceptable univariate indicators, we acknowledged that the 5-point Likert scale data are fundamentally ordinal and often deviate from the stricter assumption of multivariate normality required for the standard Maximum Likelihood (ML) estimator. Therefore, to ensure the robustness of the results, the Maximum Likelihood with Robust standard errors (MLR) estimator was selected. This estimator provides parameter estimates that are robust to non-normality and computes a scaled chi-square test statistic and robust standard errors, yielding more accurate model fit indices and significance tests (Kline, 2011). Model fit was evaluated using multiple indices and established cut-off criteria: the ratio of chi-square to degrees of freedom (χ^2 / df) < 3, Comparative Fit Index (CFI) > .95, Tucker-Lewis Index (TLI) > .95, Root Mean Square Error of Approximation

(RMSEA) < .08, and Standardized Root Mean Square Residual (SRMR) < .08 (Hair *et al*., 2019; Kline, 2011; Marsh *et al*., 1988).

For criterion and discriminant validity, relationships between variables were assessed using Pearson correlation and two separate linear regression analyses. Prior to each regression, key statistical assumptions were thoroughly verified. For the first regression model predicting Willingness to Communicate (WTC), multicollinearity was not a concern, as VIF values were 1.51 and Tolerance values were .660. The independence of residuals was confirmed with a Durbin-Watson test statistic of 2.00. Homoscedasticity was supported by a non-significant Breusch-Pagan test (p = .878). The normality of residuals was supported by a non-significant Kolmogorov-Smirnov test (p = .144). Finally, influential cases were examined using Cook's distance. The maximum Cook's distance was .0732, indicating that while some cases had influence, no case was excessively influential (i.e., Cook's distance > 1). A supplementary analysis showed no substantive changes in the results upon removal of the most influential case; therefore, all cases were retained. For the second regression model predicting exam scores, multicollinearity was not an issue, with all VIF values below 1.75 and Tolerance values above .570. The Durbin-Watson statistic was 1.74, suggesting no significant autocorrelation. The Breusch-Pagan test was non-significant (p = .166), indicating that the assumption of homoscedasticity was met. The normality of residuals was supported by a non-significant Kolmogorov-Smirnov test (p = .446). Influential cases were also assessed via Cook's distance, with a maximum value of .0401, suggesting no single case unduly influenced the model's parameters.

To examine measurement invariance, a multi-group confirmatory factor analysis (MG-CFA) was conducted across language proficiency levels (B1, B2, C1) and genders. Invariance was tested sequentially at the configural, metric, scalar, strict, and structural levels. The establishment of invariance at each level was determined by evaluating the change in model fit, primarily using the change in CFI (ΔCFI). A change of less than .01 (ΔCFI < .01) was considered supportive of invariance at that level (Chen, 2007).

For reliability analysis, both Cronbach's alpha (α) and McDonald's omega (ω) were calculated to provide a comprehensive assessment of internal consistency. While Cronbach's alpha is reported for conventional purposes and comparability, with values above .60 considered acceptable (Taber, 2018; van Griethuijsen *et al*., 2015), McDonald's omega was also included as it provides a more robust estimate of reliability that is not dependent on the strict assumption of tau-equivalence (i.e., that all items have equal factor loadings). Additionally, the average inter-item correlation (AIC) was examined to evaluate item homogeneity.

RESULTS

Language Validity

To assess linguistic equivalence following the back-translation process, a study was conducted with 18 bilingual participants (Sample 1), all self-reporting at least B1 proficiency in both languages. For this study, the original scale items were contextually adapted by replacing the word "English" with "Turkish". Participants completed the English form first, followed by the Turkish form after a two-week interval. The total scores from the two administrations were compared using the Wilcoxon signed-rank test, as the data violated the assumption of normality (Shapiro-Wilk, p < .05). The analysis revealed no significant difference between the total scores of the English and Turkish forms (Z = 0.188, p = .872), confirming the linguistic equivalence of the adapted scale.

Construct Validity

First, we conduct item analysis using the data from sample 2 (n = 234). Item-test correlations of the L2-Grit scale vary between .42 and .58. This is above the cut-off value of .30 and means that the items of the measurement tool are consistent with the total structure (Field, 2017). We

also found the reliability values of the scale as .81 overall, .84 in Factor 1, and .81 in Factor 2. Thus, α reliability values of the scale and sub-factors are sufficient (Field, 2017). After, for construct validity, we conduct confirmatory factor analysis. According to CFA, factor loadings vary between .55 and .86. It is stated that the minimum value should be .50 (Hair *et al.*, 2019). The fit indices for the hypothesis model were SRMR = .04, RMSEA = .07 (95% CI [.04, .09]), CFI = .96, TLI = .94, χ^2 / *df* = 2.27. For the second order factor, the fit indices were SRMR = .04, RMSEA = .07 (95% CI [.05, .10]), CFI = .96, TLI = .94, χ^2 / *df* = 2.36. According to the information obtained, it can be said that both models have values within acceptable limits (Hair *et al.*, 2019; Kline, 2011; Marsh *et al.*, 1988). Figure 1 shows the path diagrams (PoE: Factor 1, CoI: Factor 2).

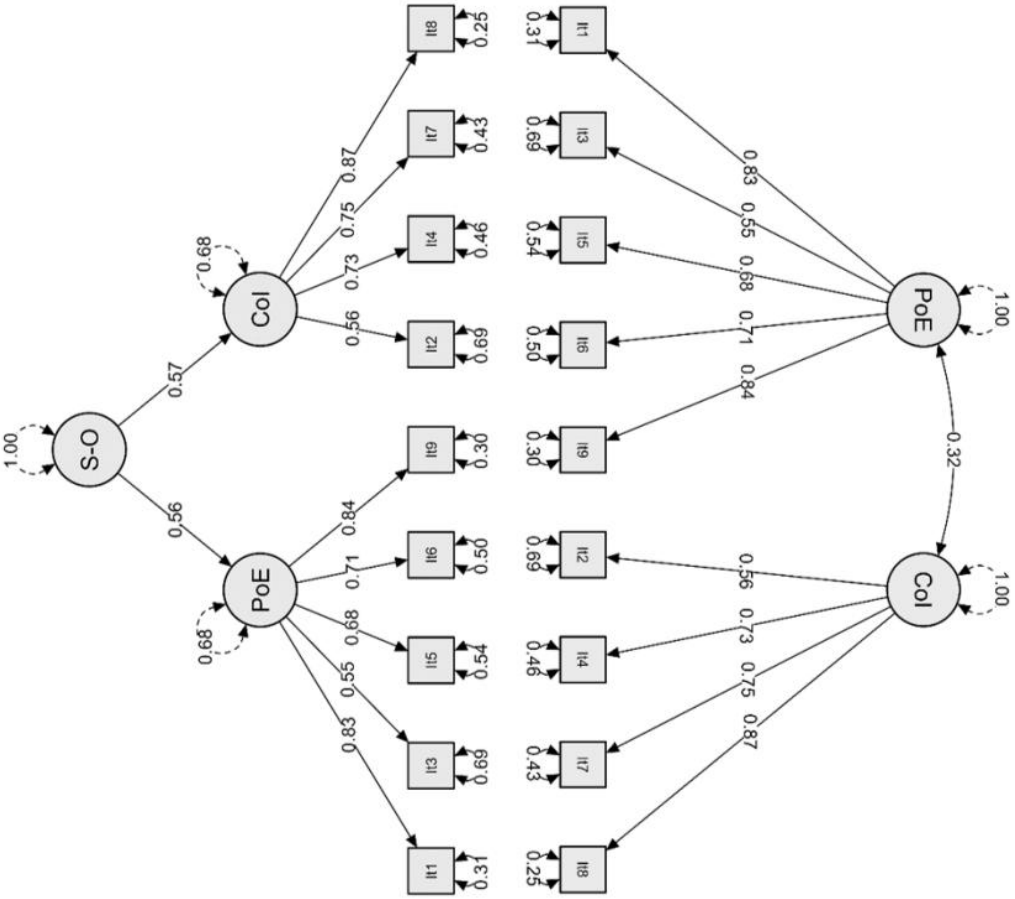

Figure 1. *CFA path diagrams.*

For further examination of construct validity, exploratory structural equation modelling (ESEM) was conducted with the merged data of Sample 3 and Sample 4 (n = 406). ESEM was employed to cross-validate the factor structure in a more flexible framework that allows for potential cross-loadings. Results from the ESEM suggested that the two-factor structure provided an acceptable model fit: χ^2 / *df* = 1.68, CFI = .97, TLI = .95, RMSEA = .04 (95% CI [.01, .07]), and SRMR = .02. Factor loadings indicated strong and significant associations for the Perseverance of Effort (PoE) factor with items GRIT1 (λ = .66), GRIT3 (λ = .50), GRIT5 (λ = .60), GRIT6 (λ = .56), and GRIT9 (λ = .56), all significant at p < .001. Similarly, the Consistency of Interest (CoI) factor was strongly loaded by items GRIT2 (λ = .54), GRIT4 (λ = .67), GRIT7 (λ = .70), and GRIT8 (λ = .79), also significant at p < .001. Correlation between factors PoE and CoI was moderate and significant (r = .31, p < .001). These findings further supported the scale's two-factor structure, demonstrating strong construct validity of the adapted L2 Grit scale. Figure 2 shows the path diagram (PoE: Perseverance of Effort, CoI: Consistency of Interest).

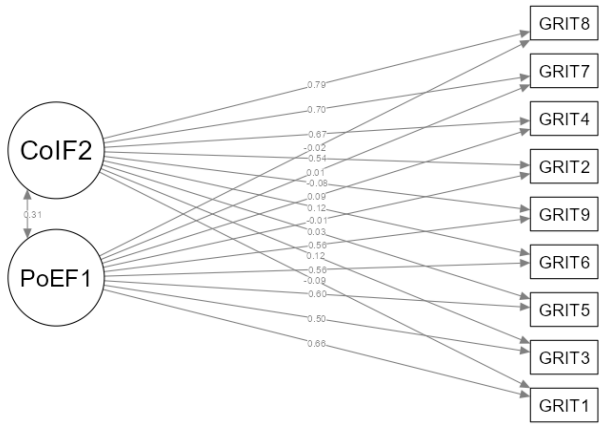

Figure 2. *ESEM path diagram.*

Convergent validity was evaluated after CFA by employing the Average Variance Extracted (AVE) and Composite Reliability (CR) measures. For this analysis we used CFA factor loadings (sample 2). CR values above .70 and AVE values above .50 are considered acceptable (Fornell & Larcker, 1981; Hair *et al.*, 2019). PoE (AVE = .53, CR = .84) and CoI (AVE = .54, CR = .83) met these criteria, confirming convergent validity.

Criterion and Discriminant Validity

For the criterion validity, we examined the correlation between L2 Grit and its sub-factors with the sample 3 (n = 186). Accordingly, we identified a significant correlation between L2 Grit and L2 Grit PoE (r = .77, p < .001) as well as L2 Grit CoI (r = .83, p < .001). A positive relationship was also observed between PoE and CoI (r = .31, p < .001). There was additionally a meaningful positive correlation between L2 Grit and general Grit (r = .58, p < .001). A regression analysis was then performed to investigate how general Grit and L2 Grit predicted willingness to communicate (WTC) in Turkish (Table 1).

Table 1. *Model coefficients – WTC.*

							Collinearity Statistics	
Model			*B*	SE	β	*t*	*p*	Tolerance	VIF
M1	(Intercept)		2.67	0.26		10.24	< .001		
	GRIT		0.29	0.07	0.28	3.96	< .001	1.00	1.00
M2	(Intercept)		2.04	0.27		7.37	< .001		
	GRIT		0.04	0.08	0.04	0.53	.591	0.66	1.51
	L2-GRIT		0.39	0.08	0.40	4.93	< .001	0.66	1.51
M1 R = .28, R^2 = .07, Adjusted R^2 = .07, F(1, 184) = 15.68, p < .001									
M2 R = .43, R^2 = .18, Adjusted R^2 = .17, F(2, 183) = 21.03, p < .001									

In the first model, general Grit significantly predicted WTC, explaining 7.4% of the variance, F(1, 184) = 15.68, p < .001, β = 0.29. When L2 Grit was introduced in the second model, the explained variance significantly increased to 17.8%, F(2, 183) = 21.03, p < .001. Notably, the previously significant predictive effect of general Grit became non-significant (β = 0.05, p = .591), whereas L2 Grit emerged as a significant and stronger predictor of WTC (β = 0.40, p < .001).

Afterward, data from sample 4 (n = 220) were collected for further correlational analyses. Results showed that L2 Grit was positively associated with exam scores (r = .27, p = .001), FLE (r = .50, p < .001), WTC in Turkish (r = .38, p < .001), and FLPOM (r = .37, p < .001). In contrast, it was inversely related to FLCA (r = -.22, p < .001). L2 Grit positively correlated with

exam performance, while anxiety showed a negative correlation (r = -.24, p < .001). FLPOM and WTC were non-significant, and FLE had a weak positive correlation (r = .14, p < .05). Regression analysis (Table 2) revealed that L2 Grit was the strongest positive predictor of students' exam scores (β = 0.25, p = .001), while classroom anxiety negatively predicted performance (β = -0.23, p = .001). FLPOM, FLE, and WTC were not significant predictors (p > .05). These results reinforce the distinct role of L2 Grit.

Table 2. *Model coefficients – exam score.*

						95% CI	
Predictor	*B*	SE	*t*	*p*	β	*LL*	*UL*
Intercept	76.41	8.17	9.35	<.001			
FLPOM	-1.26	1.46	-0.86	.388	-0.06	-.22	.08
FLE	0.87	1.57	0.55	.577	0.04	-.12	.21
WTC	-1.24	1.38	-0.89	.370	-0.07	-.22	.08
FLCA	-3.61	1.12	-3.23	.001	-0.23	-.37	-.09
L2-GRIT	5.05	1.55	3.26	.001	0.25	.09	.40
R = .34, R^2 = .11, F(5, 214) = 5.73, p < .001							

Measurement Invariance and Reliability

By combining the three samples from construct, criterion, and discriminant validity analyses, we obtained a final dataset of 640 participants. This allowed a comprehensive evaluation of the L2 GRIT scale's reliability and measurement invariance. Measurement invariance was tested via multi-group CFA across proficiency levels (B1, B2, C1) and gender, examining configural, metric, scalar, strict, and structural invariance (Table 3).

Table 3. *Measurement invariance across different proficiency levels and genders.*

		χ^2	*df*	CFI	TLI	RMSEA	SRMR
Level	Configural	143.284	78	.959	.944	.063	.053
	Metric	158.606	92	.959	.951	.058	.058
	Scalar	176.652	106	.956	.955	.056	.056
	Strict	203.434	124	.951	.957	.055	.059
	Structural	213.954	134	.950	.960	.053	.070
Gender	Configural	96.383	52	.972	.961	.052	.040
	Metric	106.927	59	.970	.963	.051	.047
	Scalar	118.141	66	.967	.964	.050	.045
	Strict	148.037	75	.954	.955	.055	.051
	Structural	153.750	80	.953	.958	.054	.056

CFA Fit Indices (n = 640): χ^2(26) = 65.751, p < .001, CFI = .975, TLI = .965, RMSEA = .049 (90% CI [.034, .064]), SRMR = .031; n_B1 = 242, n_B2: 178, n_C1 = 220

CFA Fit Indices (n = 636, 4 participants did not reveal their gender): χ^2(26) = 66.252, p < .001, CFI = .974, TLI = .964, RMSEA = .049 (90% CI [.034, .064]), SRMR = .034; n_female = 296, n_male = 340

Configural invariance tests showed that the factor structure was stable across groups, with fit indices indicating an adequate model fit to the data. These findings suggest that the basic structure remained consistent across proficiency levels. Metric invariance tests demonstrated equivalence of factor loadings across groups. The minor changes in CFI (ΔCFI = -.007) and RMSEA (ΔRMSEA = -.005) were within acceptable limits, supporting metric invariance. Scalar invariance tests further indicated that item intercepts were invariant across groups. The observed ΔCFI (-.003) and ΔRMSEA (-.002) confirmed support for scalar invariance. Strict invariance tests whether residual variances are equivalent across groups. Although CFI decreased slightly (ΔCFI = -.005), it remains within the acceptable range, supporting strict

invariance. Structural invariance examines whether factor variances and covariances are equivalent. The observed ΔCFI (-.001) and ΔRMSEA (-.002) indicate support for strict invariance.

When testing gender groups, configural invariance results showed a good fit to the data. This indicated that the factor structure was consistent across gender. Metric invariance by gender was supported as factor loadings were invariant across groups. ΔCFI (-.002) and ΔRMSEA (-.001) values suggested adequate support for metric invariance. Scalar invariance analysis confirmed intercept equivalence across genders. The change in CFI (ΔCFI = -.003) and RMSEA (ΔRMSEA = -.001) falls within acceptable thresholds, indicating that scalar invariance holds. were. Although CFI decreased slightly (ΔCFI = -.013), strict invariance results remain within the acceptable range, supporting strict invariance. Structural invariance was also assessed. The small changes in ΔCFI (-.001) and ΔRMSEA (-.001) confirmed structural invariance (Chen, 2007).

The L2-GRIT scale demonstrated positive correlations with PoE-F1 (r = .81, p < .001) and Col-F2 (r = .79, p < .001), indicating substantial associations between these constructs. Additionally, PoE-F1 and Col-F2 were moderately correlated (r = .29, p < .001), suggesting a meaningful but weaker relationship compared to their correlations with L2-GRIT. The internal consistency of the L2-GRIT scale was assessed using α and ω. The results indicate that both estimates were acceptable to good, with α = .76 (95% CI [.74, .79]) and ω = .77 (95% CI [.74, .80]), demonstrating reliable internal consistency. Consistent with these results, Guttman's λ2 was .79 (95% CI [.76, .80]), slightly exceeding α (.77) and ω (.76), which reinforces the scale's reliability. The average inter-item correlation (AIC = .27) falls within an acceptable range, suggesting a balanced level of homogeneity among the items. Furthermore, item-total correlations ranged between .392 and .541, all exceeding the commonly recommended .30 cutoff supporting the contribution of each item to the overall construct (Field, 2017).

DISCUSSION and CONCLUSION

Discussion

In this study, we carried out an adaptation of the L2 Grit scale within Turkish language instruction, producing a reliable and valid version of the scale. Previous research on the construct validity of the L2 Grit scale has been performed in diverse educational contexts, including Farsi students (Teimouri *et al.*, 2020), Chinese EFL students (Wei *et al.*, 2020), Polish EFL students (Pawlak *et al.*, 2022), and international university students from Canada, Japan, South Korea, and the United States (Sudina *et al.*, 2021) in the context of English instruction. Also, in the German educational context both general and domain-specific forms of grit were validated (Schmidt *et al.*, 2019) and investigated "wording-effect", resulting that the effect can occur, but in the meantime, still separation of goals can be observed (Rost *et al.*, 2024). Additionally, an adaptation study has been conducted with EFL students in Türkiye (Uştuk & Erarslan, 2023). However, no study has addressed grit in the setting of teaching Turkish as a foreign language.

The first phase of our scale adaptation process focused on language validity, using a back-translation approach. We administered both forms to bilingual individuals with intervals, finding no significant differences between the two measurements, thus completing the language validity phase. Uştuk & Erarslan (2023) implemented a similar process in their translation of the L2 Grit scale within an EFL context.

To assess construct validity, item analysis, CFA, and convergent validity tests were conducted on Sample 2 (n = 234), followed by ESEM on the merged Sample 3 and Sample 4 dataset (n = 406). CFA outcomes supported a two-factor solution, showing strong model fit and significant loadings. Convergent validity was supported, as CR and AVE values met the recommended thresholds in Sample 2 CFA: PoE (AVE = .53, CR = .84) and CoI (AVE = .54, CR = .83)

(Fornell & Larcker, 1981; Hair *et al.*, 2019). ESEM further substantiated the validity of the scale by showing high fit indices. It also revealed a moderate correlation between PoE and CoI (r = .31), which is consistent with previous findings (Pawlak *et al.*, 2022; Wei *et al.*, 2020).

In assessing criterion validity, this study intentionally selected variables that would both replicate the core findings of the original validation study and extend the construct's nomological network. Similar to Teimouri *et al*. (2020), we examined the relationship between L2 Grit and key variables such as Willingness to Communicate (WTC), language anxiety (FLCA), language enjoyment (FLE), and academic achievement (exam scores). The positive and significant associations found in our study with these constructs not only confirm the scale's validity in the Turkish context but also replicate the foundational findings of the original research, strengthening the cross-cultural relevance of L2 grit. Furthermore, to deepen the understanding of L2 grit's role among affective variables, this study also included a measure of Foreign Language Peace of Mind (FLPOM), a construct not present in the original study. Therefore, the criterion validity analyses presented here serve a dual purpose: they confirm the core predictive power of L2 grit while also providing new evidence of its relationship with other positive emotions in language learning. A moderate and significant correlation was found between Turkish context-specific L2 Grit and domain-general Grit, which is consistent with findings from prior research (Alrabai, 2024; Demir, 2024). Hierarchical regression analysis showed that L2 Grit predicted WTC in Turkish more strongly than general grit, reinforcing both discriminant and criterion validity. Prior research also links L2 grit and WTC (Alrabai, 2024; Bensalem *et al*., 2023; Cheng, 2021; Lee, 2022). Individuals who feel confident and competent in foreign language learning are likelier to initiate communication (Aoyama & Takahashi, 2020; Fallah, 2014; Ghanbarpour, 2016). From an affective perspective on grit, individuals who feel self-assured and capable in their target language exhibit higher levels of grit.

Additionally, L2 Grit showed significant positive correlations with FLE, FLPOM, and WTC, and a negative correlation with FLCA. These results are similar with previous research highlighting positive associations between grit and enjoyment and negative associations with anxiety. Pawlak *et al.* (2022) found correlations between language learning enjoyment (r = .53) and foreign language anxiety (r = -.26). Wei *et al.* (2020) further noted that "both 'L2 grit' and 'L2 joy' are positive psychological variables, which tend to correlate intrinsically." In our adaptation, we found a positive correlation between L2 grit and L2 enjoyment (r = .50). Other studies have likewise reported that enjoyment is positively linked with grit, whereas anxiety tends to show a negative relationship with grit. Which aligns with the nature of grit as it requires sustained effort and perseverance. Highly anxious individuals often lack the stability necessary for the long-term resilience that grit demands (Mercer & Ryan, 2010; Teimouri *et al.*, 2019; Zarrinabadi & Pawlak, 2021). Furthermore, if individuals do not enjoy their learning activity, it can become a source of stress or even "suffering." Thus, enjoyment in language learning is a necessary and effective factor for sustaining grit (Fathi & Hejazi, 2024).

Regarding predictive validity, regression analyses confirmed that L2 Grit significantly predicts academic achievement (exam performance), reinforcing the distinct and critical role of domain-specific perseverance. Conversely, anxiety negatively predicted exam performance. This result aligns with previous research indicating that domain-specific grit fosters L2 proficiency, long-term engagement, and resilience in second language acquisition (Alrabai, 2024; Pawlak *et al.*, 2022; Teimouri *et al.*, 2020; Wei *et al.*, 2020).

Measurement invariance analyses confirmed configural, metric, scalar, strict, and structural invariance across different proficiency levels (B1, B2, C1) and genders, demonstrating the scale's consistent measurement properties. Since configural, metric, scalar, strict, and structural invariance were all supported, it can be concluded that the L2 Grit scale measures the same construct equivalently across proficiency levels and genders. The establishment of scalar invariance allows for valid comparisons of latent means across B1, B2, and C1 groups. Furthermore, the results of strict and structural invariance suggest that measurement precision

and construct relationships remain stable, reinforcing the scale's robustness for assessing L2 Grit in different proficiency levels.

Regarding the reliability of the adapted scale, the findings of this study confirmed strong internal consistency, with a Cronbach's alpha of .76 and a McDonald's omega of .77 for the total scale. These values are comparable to those reported in the original validation study by Teimouri *et al*. (2020) and other adaptations, such as the Turkish EFL context validation by Uştuk & Erarslan (2023). This consistency across different linguistic and cultural contexts further supports the robustness of the L2 Grit construct and the reliability of its measurement with this scale.

Conclusion

This study adapted and validated the L2 Grit scale for Turkish as a foreign language, providing a reliable instrument for assessing perseverance and consistency of interest among language learners (Appendix 1). Through rigorous validation processes, including language validity, CFA, ESEM, and criterion analyses, the adapted scale showed strong psychometric properties. Significant positive relationships emerged between L2 grit and motivational variables such as enjoyment and willingness to communicate, while negative correlations with language anxiety highlighted grit's protective role in language learning contexts. Additionally, regression analysis confirmed that L2 Grit is a stronger predictor of Willingness to Communicate (WTC) in Turkish compared to general Grit, underscoring the importance of domain-specific perseverance. Measurement invariance tests demonstrated the scale's consistent measurement properties across proficiency levels (B1, B2, C1) and genders, enabling meaningful comparisons across learner groups. These findings collectively illustrate that fostering L2-specific grit enhances learners' motivation, reduces anxiety, and supports sustained language engagement. The validated L2 Grit scale thus serves as a robust, culturally relevant tool, offering valuable applications for researchers and educators aiming to support learners' long-term perseverance and success in Turkish language acquisition.

Practical Implications

The findings of this study offer several actionable recommendations for practitioners in the field of teaching Turkish as a foreign language:

Diagnostic and Reflective Tool for Educators: Teachers can use the Turkish L2 Grit Scale at the beginning of an academic term to identify students with lower levels of perseverance or interest. This early diagnosis can help in providing targeted motivational support to students who may be at risk of disengagement or dropout. Additionally, the scale can serve as a reflective tool for teachers to assess whether their classroom environment fosters the necessary persistence for long-term language learning.

Creating a Supportive Affective Climate: The study found a strong positive correlation between L2 grit and foreign language enjoyment (FLE) (r = .50) and a negative correlation with foreign language classroom anxiety (FLCA) (r = -.22). This suggests that grit is more likely to flourish in a positive and low-stress environment. Therefore, a key pedagogical implication is for teachers to focus on creating a supportive, enjoyable, and psychologically safe classroom climate. Strategies such as reducing the fear of making mistakes, increasing collaborative and engaging activities, and praising effort over innate talent can mitigate anxiety and boost enjoyment, thereby creating fertile ground for the development of grit.

Fostering Long-Term Interest and Effort: To address both dimensions of grit, curriculum designers and teachers can integrate long-term, challenging projects that require sustained effort over several weeks or a semester. Furthermore, to cultivate "Consistency of Interest," educators can help students connect their personal long-term goals (e.g., studying in Türkiye, connecting with relatives) to the daily tasks of language learning. This helps students see the relevance of their efforts and maintain their passion over time.

Promoting Willingness to Communicate (WTC): Our findings revealed that L2 grit was a stronger predictor of WTC than general grit. This is a powerful practical insight: by fostering domain-specific grit through the strategies above (i.e., increasing enjoyment, reducing anxiety, and connecting to long-term goals), teachers can directly encourage students to participate more actively and willingly in classroom communication.

Limitations and Future Directions

Despite its contributions, this study has several limitations that should be addressed in future research. Firstly, the use of convenience sampling from Turkish state universities limits the generalizability of findings. Future studies should aim to include more diverse samples from different instructional settings (e.g., private language courses, learners in other countries). Secondly, the study's cross-sectional design does not allow for causal inferences. Longitudinal research is needed to track the development of L2 grit over time and its causal relationship with achievement. Thirdly, both grit and language proficiency were measured via self-report. Future research could incorporate more objective measures, such as teacher ratings of grit or standardized proficiency tests, to provide a more comprehensive picture. Lastly, this study did not explore the cultural factors that might influence grit. Cross-cultural comparative studies could offer valuable insights into how L2 grit manifests and develops in different learning contexts).

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Ordu University, Education Research Ethics Committee, 2024-127.

Data Availability Statement

Research data is available and can be sent with e-mail upon reasonable request.

Informed Consent

Written informed consent was obtained from the participants who participated in this study.

Contribution of Authors

E.Ç.: Investigation, Methodology, Software, Formal analysis, and Writing-original draft. **M.E.U.**: Literature review, Investigation, Methodology, Writing-original draft.

Orcid

Enes Çinpolat https://orcid.org/0000-0002-3411-4300
Muhammed Eren Uygur https://orcid.org/0000-0003-3740-7579

REFERENCES

Akın, A., & Arıcı, N. (2015). Kısa kararlılık ölçeğinin Türkçe formunun (KKÖ): Geçerlik ve güvenirlik çalışması [Turkish version of the short grit scale: The study of validity and reliability]. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi*, *52*, 256-264. https://dergipark.org.tr/tr/pub/abuhsbd/issue/32946/366109

Alamer, A. (2021). Grit and language learning: Construct validation of L2-Grit scale and its relation to later vocabulary knowledge. *Educational Psychology*, *41*(5), 544-562. https://doi.org/10.1080/01443410.2020.1867076

Alrabai, F. (2024). Modeling the relationship between classroom emotions, motivation, and learner willingness to communicate in EFL: Applying a holistic approach of positive psychology in SLA research. *Journal of Multilingual and Multicultural Development*, *45*(7), 2465-2483 https://www.tandfonline.com/doi/abs/10.1080/01434632.2022.2053138

Aoyama, T., & Takahashi, T. (2020). International students' willingness to communicate in English as a second language: The effects of l2 self-confidence, acculturation, and motivational types. *Journal of International Students*, *10*(3), 703-723. https://doi.org/10.32674/jis.v10i3.730

Ateş, A. (2019). The impact of the emotional intelligence of learners of Turkish as a foreign language on reading comprehension skills and reading anxiety. *Universal Journal of Educational Research*, *7*(2), 571-579. https://doi.org/10.13189/ujer.2019.070230

Bensalem, E., Thompson, A.S., & Alenazi, F. (2023). The role of grit and enjoyment in EFL learners' willingness to communicate in Saudi Arabia and Morocco: A cross-cultural study. *Journal of Multilingual and Multicultural Development*, *46*(3), 749-764. https://www.tandfonline.com/doi/abs/10.1080/01434632.2023.2200750

Bi, J., Izadpanah, S., Mohammadi, Z., & Rezaei, Y.M. (2024). Investigating the impact of technology-based education on academic motivation, academic perseverance, and academic self-efficacy in English language learning skills. *Education and Information Technologies*, *29*, 20523-20545. https://doi.org/10.1007/s10639-024-12712-0

Botes, E., Azari Noughabi, M., Amirian, S.M.R., & Greiff, S. (2024). New wine in new bottles? L2 Grit in comparison to domain-general Grit, Conscientiousness, and Cognitive Ability as a predictor of language learning. *Journal of Multilingual and Multicultural Development*, *46*(8), 2451-2466. https://doi.org/10.1080/01434632.2023.2294120

Botes, E., Dewaele, J.M., & Greiff, S. (2021). The development and validation of the short form of the foreign language enjoyment scale. *The Modern Language Journal*, *105*(4), 858–876. https://doi.org/10.1111/modl.12741

Botes, E., Van der Westhuizen, L., Dewaele, J.M., MacIntyre, P., & Greiff, S. (2022). Validating the short-form foreign language classroom anxiety scale. *Applied Linguistics*, *43*(5), 1006-1033. https://doi.org/10.1093/applin/amac018

Brislin, R.W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, *1*(3), 185-216. https://doi.org/10.1177/135910457000100301

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504. https://doi.org/10.1080/10705510701301834

Cheng, H.F. (2021). The effect of language-specific grit and future self-guides on willingness to communicate in the second language. *English Teaching & Learning*, *45*(3), 283-301. https://doi.org/10.1007/s42321-021-00081-7

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

Çinpolat, E., Mutlu, H.H., & Uygur, M.E. (2024). Türkçenin yabancı dil olarak öğretimi bağlamında iletişim kurma istekliliği: Ölçek uyarlama çalışması [Willingness to communicate in the context of teaching Turkish as a foreign language: A scale adaptation study]. *Türk Eğitim Bilimleri Dergisi, 22*(3), 1954-1972. https://doi.org/10.37217/tebd.1547814

Demir, Y. (2024). L2 grit: A structured approach to preliminary biblio-systematic review. *System*, *123*, Article 103353. https://doi.org/10.1016/j.system.2024.103353

Dewaele, J.M., & MacIntyre, P. (2016). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, *4*(2), 237-274. http://dx.doi.org/10.14746/ssllt.2014.4.2.5

Duckworth, A.L. (2017). *Grit: Why passion and resilience are the secrets to success*. Random House.

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., 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

Evans, M., & Tragant, E. (2020). Demotivation and dropout in adult EFL learners. *The Electronic Journal for English as a Second Language*, *23*(4), 1-20. https://eric.ed.gov/?id=EJ1242642

Fallah, N. (2014). Willingness to communicate in English, communication self-confidence, motivation, shyness and teacher immediacy among Iranian English-major undergraduates: A structural equation modeling approach. *Learning and Individual Differences*, *30*, 140-147. https://doi.org/10.1016/j.lindif.2013.12.006

Fathi, J., & Hejazi, S.Y. (2024). Ideal L2 self and foreign language achievement: The mediating roles of L2 grit and foreign language enjoyment. *Current Psychology*, *43*(12), 10606-10620. https://doi.org/10.1007/s12144-023-05187-8

Fathi, J., Pawlak, M., Saeedian, S., & Ghaderi, A. (2024). Exploring factors affecting foreign language achievement: The role of growth mindset, self-efficacy, and L2 grit. *Language Teaching Research*, 1-32. https://doi.org/10.1177/13621688241227603

Field, A. (2017). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications Ltd.

Fornell, C., & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.2307/3151312

Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. (2022). *How to design and evaluate research in education* (11th ed.). McGraw Hill.

Ghanbarpour, M. (2016). Willingness to communicate, linguistic self-confidence, and language-use anxiety: The Iranian EFL context. *Theory and Practice in Language Studies*, *6*(12), 2265-2271. https://doi.org/10.17507/tpls.0612.05

Gurbuz, H. (2022). Learning and learners' inability to learn a foreign language. *International Journal of Social Science Research and Review*, *5*(9), 323-328. https://doi.org/10.47814/ijssrr.v5i9.442

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2019). *Multivariate data analysis*. Cengage.

Hambleton, R. K., & Patsula, L. (1999). Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1*, 1-12. https://www.testpublishers.org/assets/documents/volume%201%20issue%201Increasing%20validity.pdf

Hambleton, R.K., Merenda, P.F., & Spielberger, C.D. (Eds.). (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Psychology Press.

Ismail, S.M., Wang, C., & Jamalyar, R. (2023). The impact of task-based instruction on learners' reading comprehension, L2 grit, anxiety, and motivation for L2 reading. *Asian-Pacific Journal of Second and Foreign Language Education*, *8*(1), Article 42. https://doi.org/10.1186/s40862-023-00216-2

Javadi Safa, A., Zarei Mavini, A.A., & Esfandiari, R. (2022). Foreign language learners dropping out of language classes: Honoring dropouts' voice (research paper). *Iranian Journal of English for Academic Purposes*, *11*(2), 38-55. https://journalscmu.sinaweb.net/article_153534.html

Kennedy, D.P. (2005). Scale adaptation and etnography. *Field Methods, 17*(4), 412-431. https://doi.org/10.1177/1525822X05280060

Khajavy, G.H., MacIntyre, P., & Hariri, J. (2020). A closer look at grit and language mindset as predictors of foreign language achievement. *Studies in Second Language Acquisition*, *43*, 379-402. https://doi.org/10.1017/S0272263120000480

Kline, R.B. (2011). *Principles and practice of structural equation modeling*. Guilford Publications.

Lee, J.S. (2022). The role of grit and classroom enjoyment in EFL learners' willingness to communicate. *Journal of Multilingual and Multicultural Development*, *43*(5), 452-468. https://doi.org/10.1080/01434632.2020.1746319

Lee, J.S., & Drajati, N.A. (2020). Willingness to communicate in digital and non-digital EFL contexts: Scale development and psychometric testing. *Computer Assisted Language Learning*, *33*(7), 688-707. https://doi.org/10.1080/09588221.2019.1588330

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*. *6*(4), 1513-1537. https://doi.org/10.1515/iral-2022-0196

Li, J., & Yuan, F. (2024). Does engagement with feedback matter? Unveiling the impact of learner engagement and grit on EFL learners' English writing achievements. *Language Teaching Research*. https://doi.org/10.1177/13621688241257865

Liu, Y. (2022). Investigating the role of English as a foreign language learners' academic motivation and language mindset in their grit: A theoretical review. *Frontiers in Psychology*, *13*, Article 872014. https://doi.org/10.3389/fpsyg.2022.872014

Marsh, H.W., Balla, J.R., & McDonald, R.P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391-410. https://doi.org/10.1037/0033-2909.103.3.391

Mercer, S., & Ryan, S. (2010). A mindset for EFL: Learners' beliefs about the role of natural talent. *ELT Journal*, *64*(4), 436-444. https://doi.org/10.1093/elt/ccp083

Mikami, H. (2024). Revalidation of the L2-Grit scale: A conceptual replication of Teimouri, Y., Plonsky, L., & Tabandeh, F. (2022). L2 grit: Passion and perseverance for second-language learning. *Language Teaching*, *57*(2), 274-289. https://doi.org/10.1017/S0261444822000544

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://psycnet.apa.org/record/2020-62514-001

Ozer, Z., & Altay, I.F. (2021). Examining the level of enjoyment and anxiety among Turkish EFL students. *Journal of Language and Linguistic Studies*, *17*(S1), 663-671. https://search.informit.org/doi/abs/10.3316/INFORMIT.167504744047536

Pawlak, M., Csizér, K., Kruk, M., & Zawodniak, J. (2022). Investigating grit in second language learning: The role of individual difference factors and background variables. *Language Teaching Research*, 13621688221105775. https://doi.org/10.1177/13621688221105775

Pawlak, M., Csizér, K., Kruk, M., & Zawodniak, J. (2022). Investigating grit in second language learning: The role of individual difference factors and background variables. *Language Teaching Research*, *29*(5), 2093-2118. https://doi.org/10.1177/13621688221105775

Rost, D.H., Greiff, S., Ladkani, S., & Stadler, M. (2024). Sticking with it: The dimensionality and domain-specificity of grit. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000837

Şakrak, G. (2009). The relationship between emotional intelligence and foreign language anxiety in Turkish EFL students [Unpublished master's thesis]. Bilkent University. http://hdl.handle.net/11693/15422

Sarıçam, H., Çelik, İ., & Oğuz, A. (2016). Kısa azim (sebat) ölçeğinin Türkçeye uyarlanması: Geçerlik ve güvenirlik çalışması [Turkish adaptation of the short grit scale (grit-s): Validity and reliability study]. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi*, *5*(2), 927-935. https://doi.org/10.7884/teke.622

Schmidt, F.T.C., Fleckenstein, J., Retelsdorf, J., Eskreis-Winkler, L., & Möller, J. (2019). Measuring Grit: A German validation and a domain-specific approach to Grit. *European Journal of Psychological Assessment*, *35*(3), 436-447. https://doi.org/10.1027/1015-5759/a000407

Sudina, E., Brown, J., Datzman, B., Oki, Y., Song, K., Cavanaugh, R., Thiruchelvam, B., & Plonsky, L. (2021). Language-specific grit: Exploring psychometric properties, predictive validity, and differences across contexts. *Innovation in Language Learning and Teaching*, *15*(4), 334-351. https://doi.org/10.1080/17501229.2020.1802468

Sun, S., & Xu, Q. (2024). *L2 grit and second/foreign language achievement: A meta-analysis* (SSRN Scholarly Paper No. 4670841). Social Science Research Network. https://doi.org/10.2139/ssrn.4670841

Taber, K.S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Taşpınar, H.K., & Külekçi, G. (2018). GRIT: An essential ingredient of success in the EFL classroom. *International Journal of Languages' Education and Teaching*, *6*(3), 208-226. https://dergipark.org.tr/en/pub/ijlet/issue/82542/1427132

Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, *41*(2), 363-387. https://doi.org/10.1017/S0272263118000311

Teimouri, Y., Plonsky, L., & Tabandeh, F. (2020). L2 grit: Passion and perseverance for second-language learning. *Language Teaching Research*, *26*(5), 893-918. https://doi.org/10.1177/1362168820921895

Uştuk, Ö., & Erarslan, A. (2023). Adaptation and initial validation of the L2-grit scale in Turkish. *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, *13*(2), 1176-1188. https://doi.org/10.30783/nevsosbilen.1276339

van Griethuijsen, R.A.L.F., van Eijck, M.W., Haste, H., den Brok, P.J., Skinner, N.C., Mansour, N., Savran Gencer, A., & BouJaoude, S. (2015). Global patterns in students' views of science and interest in science. *Research in Science Education*, *45*(4), 581-603. https://doi.org/10.1007/s11165-014-9438-6

Wang, L. (2021). The role of students' self-regulated learning, grit, and resilience in second language learning. *Frontiers in Psychology*, *12*, Article 800488. https://doi.org/10.3389/fpsyg.2021.800488

Wei, R., Liu, H., & Wang, S. (2020). Exploring L2 grit in the Chinese EFL context. *System*, *93*, Article 102295. https://doi.org/10.1016/j.system.2020.102295

Xu, L., Wang, Z., Tao, Z., & Yu, C. (2022). English-learning stress and performance in Chinese college students: A serial mediation model of academic anxiety and academic burnout and the protective effect of grit. *Frontiers in Psychology*, *13*, Article 1032675. https://doi.org/10.3389/fpsyg.2022.1032675

Yang, Y., Cui, Y., & Yao, S. (2024). Teacher support, grit and L2 willingness to communicate: The mediating effect of foreign language enjoyment. *BMC Psychology*, *12*(1), Article 383. https://doi.org/10.1186/s40359-024-01877-5

Yeşilçınar, S., & Erdemir, N. (2023). Are enjoyment and anxiety specific to culture? An investigation into the sources of Turkish EFL learners' foreign language enjoyment and anxiety. *Innovation in Language Learning and Teaching*, *17*(2), 453-468. https://doi.org/10.1080/17501229.2022.2063295

Yeşilçınar, S., Dewaele, J.M., & Erdemir, N. (2025). The psychometric validation of Turkish versions of four scales: The foreign language enjoyment, peace of mind, boredom, and classroom anxiety scales. *Perceptual and Motor Skills*, *132*(1), 61-92. https://doi.org/10.1177/00315125241280333

Zarrinabadi, N., & Pawlak, M. (Eds.). (2021). *New perspectives on willingness to communicate in a second language*. Springer International Publishing. https://doi.org/10.1007/978-3-030-67634-6

Zhao, X., & Wang, D. (2023). Grit in second language acquisition: A systematic review from 2017 to 2022. *Frontiers in Psychology*, *14*, Article 1238788. https://doi.org/10.3389/fpsyg.2023.1238788

Zhou L., Dewaele J.M., Lochtman K., & Xi Y. (2023). Foreign language peace of mind: A positive emotion drawn from the Chinese EFL learning context. *Applied Linguistics Review*, *14*(5), 1385-1410. https://doi.org/10.1515/applirev-2021-0080

APPENDIX

Appendix 1. *L2 Grit in the teaching Turkish as a foreign language context.*

1. Ben gayretli bir Türkçe öğrencisiyim.	1	2	3	4	5
2. Türkçe öğrenmeye olan ilgim yıldan yıla/aydan aya değişir.	1	2	3	4	5
3. Türkçe söz konusu olduğunda, çalışkan bir öğrenciyim.	1	2	3	4	5
4. Türkçe öğrenmeye olan ilgimi kaybettiğimi düşünüyorum.	1	2	3	4	5
5. Türkçe öğrenmeye karar verdiğim için artık hiçbir şey bu hedefime ulaşmamı engelleyemez.	1	2	3	4	5
6. Hiçbir şeyin Türkçe öğrenmede beni engellemesine izin vermem.	1	2	3	4	5
7. Türkçe öğrenmeye eskisi kadar ilgi duymuyorum.	1	2	3	4	5
8. Geçmişte Türkçe öğrenme konusunda takıntılıydım ancak daha sonraları ilgimi kaybettim.	1	2	3	4	5
9. Türkçe öğrenme konusundaki zayıf yönlerimi geliştirmek için çok zaman ve çaba harcarım.	1	2	3	4	5

Not. 1, 3, 5, 6 ve 9. maddeler Çabayı Sürdürme [Perseverance of Effort] boyutuna aittir. Bu maddeler için puanları şu şekilde verin: 5 = Bana çok benziyor; 4 = Bana oldukça benziyor; 3 = Bana biraz benziyor; 2 = Bana pek benzemiyor; 1 = Bana hiç benzemiyor. 2, 4, 7 ve 8. maddeler ise İlgi Tutarlığı [Consistency of Interest] boyutunun bir parçasıdır. Bu maddeler için ters puanlama kullanın: 5 = Bana hiç benzemiyor; 4 = Bana pek benzemiyor; 3 = Bana biraz benziyor; 2 = Bana oldukça benziyor; 1 = Bana çok benziyor. Puanlama yaptıktan sonra tüm yanıtları toplayın ve toplamı 9'a bölün. Elde edilen ortalama 1 ile 5 arasında olacaktır; 5 çok yüksek düzeyde D2 direşmeyi yansıtırken 1 çok düşük düzeyde D2 direşmeyi gösterir.

[*Note*. Items 1, 3, 5, 6, and 9 belong to the Perseverance of Effort dimension. For these items, assign scores as follows: 5 = Very much like me; 4 = Mostly like me; 3 = Somehow like me; 2 = Not much like me; 1 = Not at all like me. Items 2, 4, 7, and 8 are part of the Consistency of Interest dimension. For these, use the reverse scoring scheme: 5 = Not at all like me; 4 = Not much like me; 3 = Somehow like me; 2 = Mostly like me; 1 = Very much like me. After scoring, add up all responses and divide the total by 9. The resulting average will be between 1 and 5, where 5 reflects very high L2 Grit and 1 indicates very low L2 Grit.]