CAMBRIDGE
UNIVERSITY PRESS

REPLICATION RESEARCH

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

Hitoshi Mikami [iD]

Chubu University, Kasugai, Japan
Email: mikami_h@isc.chubu.ac.jp

Abstract

This study is a conceptual replication of Teimouri et al.'s (2022) investigation into the validity of the second language (L2) grit scale (the L2-Grit scale). There are several concerns about the generalizability of the findings of Teimouri et al. (2022), especially regarding the discriminant validity of the scale and the relation of L2 grit with language achievements. A conceptual replication study was conducted because these concerns could be addressed by using a different methodology. The main findings include: (a) the factor structure of L2 grit was supported in the replication sample (106 English majors at a Japanese university), (b) the results support the discriminant validity of L2 grit, but in a different way from the initial study, and (c) L2 grit was a consistent predictor of L2-specific Grade Point Average and standardized test score. The results obtained lend further support for the validity of the L2-Grit scale.

1. Introduction

Grit is a personality trait defined as 'perseverance and passion for long-term goals' (Duckworth et al., 2007, p. 1087), which is also understood as the mental stamina necessary to pursue long-term goals despite challenges and obstacles. Early evidence suggested that grit uniquely predicted overall academic performance, such as Grade Point Average (GPA), final education level, and high-school graduation rate (Duckworth et al., 2007; Eskreis-Winkler et al., 2014). Findings like these have recently received widespread attention among researchers in the field of second language (L2) acquisition (SLA), given the mental stamina that may be required for the long and continuous process of L2 development. Indeed, although the research on grit in the field of SLA is still in its infancy, a number of relevant studies have already been published (Ebadi et al., 2018; Feng & Papi, 2020; Khajavy et al., 2021; Sudina et al., 2021; Teimouri et al., 2022; Wei et al., 2020).

As reviewed in Teimouri et al. (2022), however, previous research findings have been inconsistent regarding the relevance of grit to language learning (see also Sudina et al., 2021). Teimouri et al. (2022) then proceeded to point out that grit in previous SLA studies has been measured by the original grit scale (Duckworth et al., 2007) or its short version (Duckworth & Quinn, 2009) – the scales that measure the broad personality trait of grit (i.e., how gritty one generally is). A possible problem with this approach is that individuals show changes in some personality characteristics across situations (Dörnyei & Ryan, 2015), and such contextual variations may also occur for grittiness. On this point, Jachimowicz et al. (2018) conducted a meta-analysis of 127 studies, and confirmed that a personal value for a particular performance domain was a necessary condition for grit to have a clear relationship with performance. This suggests variability in the personality trait of grit, and provides a reason to consider learners' personality in language learning/use contexts in exploring how grit is related to SLA.

Check for updates

For this reason, Teimouri et al. (2022) developed a language-domain-specific measure of personality named the L2-Grit scale, and examined its value in SLA research. As will be discussed in detail, their results suggested that L2 grit is a unique construct from other well-known personality traits, and that the new scale is a more sensitive tool than Duckworth et al.'s (2007) grit scale for examining how personality traits are related to language learning.

It seems that the L2-Grit scale has the potential to advance our understanding of grit in language learning. This paper, however, argues that there are several limitations in the validation of this new scale. The first is high measurement error in the criterion personality measures used for assessing the construct validity of L2 grit. Because measurement error attenuates relations among variables, a question remains about the uniqueness of L2 grit from other personality traits. The second is that while Teimouri et al. (2022) showed that L2 grit was related to performance in several L2 classes and self-assessed L2 proficiency, it remains unknown how L2 grit is related to overall performance in L2 courses and objective L2 proficiency. Lack of such data limits our knowledge of the relationship between grittiness and linguistic achievement. These limitations could be overcome through the use of better-validated criterion measures of personality, L2 course-only GPA, and standardized language testing. This study therefore aimed to provide a conceptual replication of Teimouri et al. (2022). In what follows, this paper provides essential information about the development and validation of the L2-Grit scale, before explaining why and how a conceptual replication should be done. It then reports the results of the replication attempt.

2. Background

2.1 L2 Grit

Previous attempts to capture the role of grit in SLA research have been unsatisfactory for two particular reasons. The first is uncertainty about the factor structure of grit. Grit was originally conceptualized as a higher order construct composed of perseverance of effort and consistency of interests (Duckworth et al., 2007). The former reflects one's persistence in the achievement of goals, and the latter the stability of interests in the pursuit of goals. On this point, Credé et al.'s (2017) meta-analysis questioned the idea that grit is characterized by the two lower-order components, based on the superiority of the perseverance component in the prediction of performance compared with either the consistency component or overall grit. A similar finding was also observed in previous L2 research. Feng and Papi (2020) found that the perseverance component of grit was related to persistence and motivational intensity in L2 learning, and both of these relationships were mediated by future L2 self-image variables; meanwhile, the consistency component of grit had no meaningful relationship with any such L2 variables, and did not even correlate with the perseverance component. Based on these results, Feng and Papi (2020) supported the distinctiveness of the two grit sub-components.

Jachimowicz et al. (2018) proposed that one solution to the problem would be to integrate the perseverance aspect of grit with personal values/preference for a performance domain, and empirically demonstrated that greater passion attainment amplifies the relationship between perseverance and performance. The implication here is that the structure of grit in SLA may be better understood by focusing more on personal values/preference that people have in learning their L2. One such attempt is Ebadi et al.'s (2018) development of a scale designed to measure Iranian learners' grit in relation to English language learning. The results of their confirmatory factor analysis (CFA) reproduced the four-factor structure suggested by a prior exploratory factor analysis: their Iranian learners' grit consisted of having (a) interests in and (b) goals for learning English, and (c) practicing a lot and (d) trying hard to learn English. Although the relative merits of using their new scale in SLA research were not empirically tested, the results of Ebadi et al. (2018) suggest that the structure of grit in the language learning/use context could differ from the originally proposed two-factor model.

The second and related point is that past studies have had inconsistent results regarding the relationship between grit and L2 achievement. Teimouri et al. (2022) attributed this to the absence of the concept of L2-specific grit, and stressed the necessity to develop a scale focusing specifically on grit in

L2 learning/use. This view was more recently supported by Khajavy et al. (2021), who explored the relationship between grit, language mindset, and final grades in a general English course, and then reached the conclusion that domain specification would be necessary to truly understand the role of grit in classroom L2 learning.

The development and validation of the L2-Grit scale was performed using data from 191 Persian speakers studying English translation at a private university. In the development of the scale, 12 items were first subjected to a principal component analysis, and a further parallel analysis was conducted. The results confirmed the presence of two main components in the scale: perseverance of effort and consistency of interest. The sub-components of the L2-Grit scale reflect one's persistence and the stability of interests specifically in L2 learning and use. Overall L2 grit accordingly represents perseverance and passion for L2 goals.

To further examine the construct validity of the new scale, Teimouri et al. (2022) evaluated the distinction between L2 grit and the Big Five personality traits (i.e., Extraversion, Conscientiousness, Neuroticism, Openness, and Emotional Stability) (Gosling et al., 2003), because grit is known to be highly correlated with its higher-order concept of Conscientiousness (Credé et al., 2017). The results of standard multiple regression analyses showed that the Big Five variables together explained only 10% of the variance in L2 grit, suggesting that the Big Five traits have limited effect on L2 grit. This finding, according to Teimouri and his colleagues, lends support to the distinction between L2 grit and the major personality traits.

Teimouri et al. (2022) then evaluated the criterion-related validity of the L2-Grit scale by comparing the behaviors of L2 grit and grit in relation to various L2 variables: three motivational variables (intended effort, willingness to communicate, and attention), fixed and growth mindsets, two emotional measures (anxiety and joy), and three achievement variables (GPA, students' grades in three English courses, and self-assessed L2 proficiency). The results of correlation analysis showed that (a) all motivational and achievement variables correlated more strongly with L2 grit than with grit, (b) the same applied to L2 joy, and (c) L2 grit had significant relationships with growth mindset and L2 anxiety. Based on these results, Teimouri et al. (2022) concluded that 'the language-domain-specific measure of grit produces clearer and more meaningful results' (p. 19). Also, Wei et al. (2020) conducted a partial replication of Teimouri et al. (2022) in a Chinese context. Their results showed that the two-factor structure of the L2-Grit scale was suitable for their sample and that L2 grit was associated with self-assessed L2 proficiency.

2.2 Justification for replication

This section will discuss the reasons why there are several concerns about the generalizability of the findings of Teimouri et al. (2022). The first point relates to the construct validity of the L2-Grit scale. In the initial study, the Big Five variables together explained 10% of the variance in L2 grit, and this result was taken as evidence for the distinction between L2 grit and the major personality traits. However, all independent variables in the initial study except for Extraversion had low reliability coefficients. The downside here is the potential attenuation effect of measurement error, which affects the estimates of (a) the relationship between L2 grit and the Big Five traits, and (b) the relative importance of each Big Five variable in explaining variance in L2 grit. As Teimouri et al. (2022) themselves noted, this problem seems to stem from the use of a very brief measure of the Big Five traits ($k = 2$ for each variable) (Gosling et al., 2003). Where this is the case, there is a good reason to use locally validated instruments for the measurement of the Big Five traits, and then re-examine the discriminant validity of the L2-Grit scale.

The second issue with Teimouri et al.'s study concerns the criterion-related validity of the new scale. Arguably, stronger evidence could be provided by employing different performance measures from those used in the initial study. As noted above, the criterion-related validity of the L2-Grit scale was partly supported by consistent correlations between L2 grit and three types of achievement measures: the English majors' GPA scores, grades in the three L2 courses, and self-assessed L2 proficiency.

The potential problem regarding the GPA score is that undergraduates, even including language majors, often need to take courses outside their discipline, which could weaken the relationship between L2 grit and GPA. On this point, Jachimowicz et al. (2018) demonstrated that perseverance, a sub-component of grit, tends to have a stronger relationship with academic measures when participants experience passion for a performance domain. This suggests that language majors' grit may not be so closely related to performance in non-L2 courses. Also, as L2 grit is a domain-specific personality trait, such a tendency may become more pronounced in the relationship between L2 grit and grades in non-L2 courses. These explain why the use of conventional GPA scores is possibly problematic for the validation of L2-specific measures. Because Teimouri et al. (2022) did not mention what percentage of their participants' GPA is relevant to L2 learning, it is worth testing whether L2 grit has a stronger relationship with overall performance in L2 courses than with overall academic performance. Here, one may say that the initial study has already shown consistent relationships between L2 grit and grades in the three L2 courses. Regarding this, Teimouri et al.'s (2022) approach has a weakness in that the target classes were selected by the researchers themselves. This selection procedure is not the best way to avoid the risk of cherry-picking, and thereby limits the generalizability of the initial findings. On this point, Khajavy et al. (2021) also suggests that grittiness may be more effective in promoting long-term learning than promoting the completion of language courses during a semester. For these reasons, the generalizability of the relationship between L2 grit and L2-specific academic performance, especially long-term performance, is worthy of investigation.

The use of a standardized proficiency test is also crucial to confirm whether L2 grit is related to the absolute levels of attainment. The relationship between grit and L2 proficiency has most often been evaluated using self-assessed proficiency scores (Feng & Papi, 2020; Sudina et al., 2021; Teimouri et al., 2022; Wei et al., 2020). This time-effective approach facilitates larger sample sizes, which have certain benefits in statistical analysis (e.g., higher statistical power). At the same time, it would be useful to know how grit is related to the scores of standardized L2 tests, because these data can be readily compared across contexts. On this point, Tomoschuk et al. (2019) reported the results of a large-scale study assessing the classification accuracy and consistency of objective and self-assessed language proficiency measures. Their conclusion was that 'for studies that need a reliable metric of language proficiency, objective measures are the better choice' (p. 535). It therefore makes sense to ask how L2 grit is related to standardized proficiency measures.

The last point of concern is the incremental validity of L2 grit. For one thing, the use of different performance measures from the initial study may change the relative importance of L2 grit compared with grit. Furthermore, what is still not clear is the merit of using L2 grit when major performance predictors other than grit are brought into the discussion. Such information is important because there have been arguments that the Big Five Conscientiousness and self-regulation traits may be better predictors of academic success than grit (Credé et al., 2017; Ivcevic & Brackett, 2014). Thus, the incremental validity of the L2-Grit scale should also be confirmed.

With these points in mind, this conceptual replication study addresses the same research questions as those of Teimouri et al. (2022):

1. How valid and reliable is the L2-Grit scale in measuring learners' perseverance and passion for L2 learning and use?
2. How is L2 grit related to language measures?

As in the initial study, this study examines the construct validity and reliability of the L2-Grit scale and its relationships to L2 achievements. Meanwhile, some changes were made in order to assess the generalizability of Teimouri et al.'s (2022) findings. As will be specified in the following sections, the assessments in this study were done using (a) a different sample, (b) different personality and performance measures, and (c) different analytical approaches from those of the initial study. All other

aspects except for the data collection procedure, which was not explicitly explained in the initial study, remain the same in this replication.

2.3 Methodology

This section will draw parallels between the methods used in this study and those used in the initial study (for side-by-side comparisons, see Table 1 at the end of this section).

2.4 Participants

One hundred and six English majors at a Japanese university participated in this study (44 female and 62 male participants, 19–22 years of age) ($Mdn = 21$, $SD = 1.05$). All participants spoke Japanese as their first language (L1), and had studied English as a school subject before enrolling in university (from six to eight years). At the time of the present investigation, the participants had learned English as their L2 on average for 2.90 years ($Mdn = 3.00$, $SD = 0.72$). The program was primarily designed to develop participants' overall English proficiency and their knowledge of English-speaking cultures. About one quarter of the credits required for their graduation could be gained in English as a second language (ESL) classes, including communicative grammar, academic writing, and integrated English. The remaining credits were to be awarded in cultural and liberal arts courses. The former included, for instance, American history and cross-cultural communication; and the latter, ecology and physical education. Note that the vast majority of the cultural and liberal arts courses were taught in the participants' L1.

The participants in this study and those in Teimouri et al. (2022) (Persian L1 participants studying English translation) have in common that (a) both Persian and Japanese have different orthography and phonology from English, (b) both sets of students had specialized in English language at the university level, and (c) their L2 learning experience and proficiency varied (see Table 1 for side-by-side comparisons). Meanwhile, it should be noted that L2 proficiency in this study was judged based on a standardized test score, while the judgement in the initial study was based on learners' self-reporting. Based on these conditions, the present sample was considered suitable for assessing the generalizability of the findings of the initial study.

The sample size of this study ($N = 106$) was smaller than that of Teimouri et al. (2022) ($N = 191$). On this point, the initial study found the average correlation of $r = .27$ between L2 grit and five performance measures. The sample size required to gain an adequate statistical power of 0.80 for $r = .27$ is $N = 105$ ($\alpha = .05$, ρ H0 = 0, 2-tailed) (Faul et al., 2007). Based on this reference value, the sample size of this study was considered acceptable for a replication attempt. It should also be noted that, owing to the concern regarding the aforementioned attenuation effect, the results of Teimouri et al.'s (2022) regression analysis were not taken into consideration in the sample size calculation.

2.5 Measurement instruments

2.5.1 Questionnaire

The 106 participants responded to a web-based questionnaire (Google Forms) written in Japanese. The questionnaire consisted of two parts. The first part asked respondents to report their name, gender, age, year at university, and language learning experience. The second part consisted of 63 items asking participants to indicate how much certain qualities apply to them on a Likert scale (from 1 'not like me at all' to 5 or 7 'very much like me'). A sample question is '私は、英語学習に勤勉である。' (i.e., 'I am a diligent English language learner') (all 63 items are available in the studies cited in this section). All questionnaire items required a response, and the order of questions was randomized for each respondent using a built-in function available in Google Forms.

Grit and L2 grit (hereafter Grit and L2 Grit) were measured using the Japanese version of the scales instead of the original English versions (the initial study). The Grit Scale ($k = 12$, the five-point scale)

was obtained from Takehashi et al. (2019) and the L2-Grit scale was downloaded from the IRIS digital repository ($k = 9$, the five-point scale) (http://www.iris-database.org). The Japanese version of the scales was used because it has previously demonstrated satisfactory reliability and validity in a Japanese population (see Table 1 for details), and the use of participants' L1 made it easier for them to respond.

The short form of the Big-Five Scale ($k = 29$, the seven-point scale) (Namikawa et al., 2012) was employed to measure the Big Five traits (i.e., Extraversion, Conscientiousness, Neuroticism, Openness, and Agreeableness). The initial study suggested the use of the Ten-Item Personality Inventory (Gosling et al., 2003) as the reason for high measurement error in the Big Five variables. Namikawa et al.'s (2012) scale appeared to be a reasonable alternative measure because it has been validated in a Japanese population.

Self-control (hereafter Self-Control), a self-regulation trait, was only documented in this study and not in the initial study. Self-Control was used for the validation owing to its close relationship with Grit (Credé et al., 2017) and academic performance (Tangney et al., 2004). The Japanese version of the Brief Self-Control Scale ($k = 13$, the five-point scale) (Ozaki et al., 2016) was used for this measurement.

2.5.2 Performance measures

With the participants' permission, their complete academic records were downloaded from the university's database. Two types of GPA scores were prepared using these records (score range: 0.00–4.00 for both measures). The first was a conventional GPA score (hereafter GPA), which was also used in the initial study. GPA was calculated using all grade-points that each participant had ever received (=total points earned/total credits attempted). The index score indicates the participants' overall academic performance. The second was named L2-GPA. L2-GPA is an alternative index to the three class grades used in the initial study, and it indicates overall performance/the average grade point in all ESL classes (the number of credits attempted: $M = 24.43$, $SD = 6.64$, $Mdn = 25.00$). The grades of English-mediated classes were excluded from this calculation because their primary goal was to develop content knowledge (e.g., the history of English-speaking countries).

The participants' L2 proficiency (hereafter Proficiency) was measured using a language test called TOEIC® instead of using self-reported information (Teimouri et al., 2022). TOEIC® is an internationally administered standardized test assessing non-native speakers' English proficiency. This test was felt to provide an appropriate measure for the present study because, being one of the most popular language tests, it will help future studies compare their results with those gained in this study. TOEIC® is a two-hour test and consists of two sections (100 comprehension questions for each section). Information on the validity and reliability of the test can be confirmed in Wei and Low (2017).

2.6 Procedure

An invitation to participate in this research project was sent via email to 150 students in the target department. The aforementioned 106 participants agreed to take part under the condition of anonymity. They also gave permission for their academic records to be accessed at this point. All participants sat the two-hour test simultaneously before completing the questionnaire over roughly 20 minutes. The procedure undertaken for this research project met the ethical requirements of the institution involved. The data collection procedure was not explicitly explained in the initial study (see Table 1 for side-by-side comparisons).

2.7 Data analysis

For statistical analysis, all negatively-worded items were reverse-scored. There were no missing data for any variable, and all variables were standardized prior to analysis. The internal consistency of the

personality measures was assessed using Cronbach's alpha. As with the initial study, the alpha values ≥.70 were judged to be acceptable. The normality of distributions was evaluated by the one-sample Kolmogorov-Smirnov (K-S) test (p should be ≥.050).

All analyses were two-tailed (N = 106). The risks of statistical Type 1 and 2 errors were set to .05 and .20, respectively. The false discovery rate method (Benjamini & Hochberg, 2000) was used to control Type I errors. Interpretation of effect sizes was done according to Ferguson's (2016) recommendations for social science data (the minimum practical effect sizes were: r ≥ .20, R^2 ≥ .04, and β ≥ .20). Ninety-five percent confidence intervals (CIs) were computed to show the precision of estimates. *SPSS* and *AMOS* (ver.27), and *G*power 3* (Faul et al., 2007) were employed for statistical computations.

There were clear differences in the validation methods used in the initial study and in the present conceptual replication. The following sections will explain the reasons for the changes (see Table 1 for side-by-side comparisons).

2.7.1 Construct validity

First, a CFA was performed to test whether the two-factor model composed of L2 Grit fit the present data. A confirmatory approach instead of an explorative method (the initial study) was used because the initial study and its replication (Wei et al., 2020) indicated support for a two-factor solution. Model fit was determined using four widely used indices: the relative chi-squared statistic (x^2/df) ≤ 3.0, the comparative fit index (CFI) and Tucker–Lewis index (TLI) ≥ .90, and the root-mean-square error of approximation (RMSEA) ≤ .08 (Harrington, 2009). The method of estimation used was maximum likelihood. It is generally recommended that the factor loading in CFA should be ≥.30 (Harrington, 2009).

Following this, a regression analysis was performed to evaluate how much variance in L2 Grit is explained by the Big Five traits (the initial study) and Self-Control. Self-Control was newly introduced to the model owing to its close relationship with Grit (Credé et al., 2017) and its relationship with academic performance (Tangney et al., 2004). All six predictors were centered to minimize multicollinearity, and then entered into the regression model. A variance inflation factor (VIF) of less than ten was taken to imply absence of multicollinearity. A stepwise approach instead of a simultaneous approach (the initial study) was used to deal with the relatively small sample size (N = 106). Backwards stepwise removal based on the likelihood-ratio statistic was used to obtain the minimum adequate model (criterion for entrance: $p < 0.05$; criterion for removal $p ≥ 0.1$).

2.7.2 Criterion-related and incremental validity

In the initial study, the Pearson's correlations were used to examine how Grit and L2 Grit were related to GPA, the grades in the three L2 classes, and self-assessed L2 proficiency. To test the generalizability of the finding that L2 Grit was associated with language performance, this study explored how the two grit variables were correlated to GPA, L2-GPA (overall performance in ESL classes), and Proficiency (the standardized test score). Conscientiousness and Self-Control (other predictors of academic success) were also added to the analysis as a preliminary step to assess the incremental validity of the L2-Grit scale. While the initial study reported only simple Pearson's correlations, this study went a step further by controlling for the influences of gender and year at university. This modification was made to deal with the possibility that these demographic factors may affect personality scores (Credé et al., 2017).

A series of hierarchical regression analyses were then conducted to test the incremental prediction of the performance scores from other variables when L2 Grit was used as the first step regression model. All predictors were centered for the analyses, and the multicollinearity was assessed based on the VIF. The incremental validity of the L2-Grit scale was not assessed in the initial study.

Table 1. Summary of the methods used in Teimouri et al. (2022) and in the present study

	Teimouri et al. (2022)	This study	Notes
Participants			
Academic level	University students		The students in both studies were specializing in English language at the university level.
L2	English		
L2 proficiency	Widely varied		Both studies found large variance in participants' L2 proficiency. Proficiency levels were self-assessed in the initial study, while proficiency in this study was assessed using a standardized test.
L1	Persian	Japanese	Both first languages have different orthography and phonology from the target language, English.
Measurements			
Questionnaire: The scales were developed by…			
L2 grit	Teimouri et al. (2022)	Sudina et al. (2021)	The Japanese version of the L2-Grit scale (Sudina et al., 2021) instead of the original English version (Teimouri et al., 2022) was used in this study, because (a) the Japanese version has previously demonstrated satisfactory reliability and factorial validity (Sudina et al., 2021), and (b) the use of participants' L1 made it easier for them to respond.
Grit	Duckworth et al. (2007)	Takehashi et al. (2019)	The Japanese version of the Grit Scale (Takehashi et al., 2019) instead of the original English version (Duckworth et al., 2007) was used in this study, because (a) the Japanese version has shown a similar level of reliability and validity as the original English version (Takehashi et al., 2019), and (b) the use of participants' L1 made it easier for them to respond.
Big Five traits	Gosling et al. (2003)	Namikawa et al. (2012)	The short form of the Big-Five Scale (Namikawa et al., 2012) instead of the Ten-Item Personality Inventory (Gosling et al., 2003) was used in this study. Namikawa et al.'s (2012) scale has been validated in a Japanese population, and the use of this locally-validated scale was expected to reduce measurement errors.
Self-Control	Not used	Ozaki et al. (2016)	Self-Control was measured only in this study. It was felt to be beneficial to include this personality trait for the validation of the L2-Grit scale, as Self-Control was reported to have a close relationship with both grit and academic performance (Tangney et al., 2004).

(*Continued*)

Table 1. (*Continued*)

	Teimouri et al. (2022)	This study	Notes
Performance: Participants' performance was represented by…			
Overall academic performance	GPA		The same index was used.
Performance in L2 classes	Grades in three L2 classes	L2 course-only GPA	The change was made to assess how personality traits were related to overall performance in ESL classes (L2 course-only GPA) rather than to performance in several L2 classes (the initial study).
L2 proficiency	Self-report	Standardized L2 test	The change was made to examine how personality traits were related to absolute levels of attainment (standardized test scores) rather than to self-assessed proficiency (the initial study).
Procedure			
Recruitment	Not reported	Via email	The initial study did not report their recruitment procedure. This study sent an email invitation to 150 students and 106 of them agreed to take part.
Access to participants' academic records	Not reported	Granted by the participants themselves	The initial study did not report how they obtained their participants' academic records (i.e., GPA and L2 course grades). In this study, the participants gave permission to access their academic records. The records were then downloaded from the database of the target university.
Proficiency test	Not used	Standardized test	The initial study did not use a test for the measurement of L2 proficiency (self-assessment was used). The participants of this study took a standardized L2 test before completing the questionnaire.
Questionnaire design and survey administration	Not reported	Web-based questionnaire following test-taking	The initial study did not report the medium of their questionnaire, the order of question items, and when their participants completed the questionnaire. The participants of this study responded to a web-based questionnaire after test-taking. The order of questions was randomized for each respondent.
Data analysis The reliability and validity of the L2-Grit scale were assessed using…			
Reliability	Cronbach's alpha		The same index was used.
Factor structure	Principle component analysis	Confirmatory factor analysis	A confirmatory approach was used in this study, because the initial study and its replication (Wei et al., 2020) indicated support for a two-factor solution.

(*Continued*)

Table 1. (*Continued*)

	Teimouri et al. (2022)	This study	Notes
Construct validity	Simultaneous multiple regression	Stepwise regression	To deal with its relatively small sample size, this study used a stepwise approach instead of a simultaneous approach (the initial study).
Criterion-related validity	Pearson's correlation	Partial correlation	The initial study used simple Pearson's correlations to examine how grit and L2 grit were related to their performance measures (GPA, the grades in the three L2 classes, and self-assessed L2 proficiency). This study tested the generalizability of the L2 grit–performance relationship by exploring how the two grit variables were related to GPA, overall performance in ESL classes, and the standardized test score. In the analysis, the influences of gender and year at university were controlled for, as such demographic factors were reported to affect personality scores (Credé et al., 2017). Other performance predictors, namely Conscientiousness and Self-Control (Ivcevic & Brackett, 2014; Credé et al., 2017), were added to the analysis as a preliminary step to assess the incremental validity of the L2-Grit scale.
Incremental validity	Not tested	Hierarchical regression	Hierarchical regression analyses were conducted only in this study. The aim of this additional testing was to assess the incremental prediction of the performance scores from other personality traits when L2 grit was used as the first step regression model.

3. Results

This section will draw parallels between the results of the present conceptual replication and those of the initial study (Porte & McManus, 2019).

3.1 Descriptive statistics

Descriptive statistics are reported in Table 2 (see online Supplementary Material A for boxplot summaries). A series of K-S tests showed that all index scores cited in Table 2 were normally distributed ($D = 0.06$–0.12, $p = .117$–$.810$) (see online Supplementary Material B-1 and B-2 for details).

The reliability of personality measures is of particular interest here because of the unacceptable Cronbach's alpha coefficients ($<.70$) observed in the initial study. The Cronbach's alpha coefficients for the Big Five variables in the current study were satisfactory (display order: this study and Teimouri et al., 2022): Extraversion ($\alpha = .89$, .79), Conscientiousness ($\alpha = .79$, .54), Neuroticism/ Emotional Stability ($\alpha = .84$, .43), Openness ($\alpha = .78$, .27), and Agreeableness ($\alpha = .73$, .11). The same applied to L2 Grit ($\alpha = .86$, .80). These results confirm that the relationship between the Big Five variables and L2 grit in this study can be assessed without great concern about the attenuation effect of measurement error.

Table 2. Descriptive statistics on the 15 indexes (raw scores)

Index	M	[95%CI]	SD	Skew	Min	Max	α
Grit	2.87	2.76, 2.98	0.57	−0.39	1.42	3.92	.76
COI	2.48	2.36, 2.61	0.67	0.20	1.00	4.17	.64
POE	3.25	3.11, 3.39	0.73	−0.40	1.50	4.67	.78
L2 Grit	3.23	3.07, 3.38	0.81	0.02	1.67	4.78	.86
L2-COI	3.50	3.31, 3.68	0.97	−0.22	1.25	5.00	.76
L2-POE	3.01	2.84, 3.18	0.90	−0.17	1.00	4.80	.89
Extraversion	4.85	4.60, 5.11	1.33	−0.29	1.00	7.00	.89
Conscientiousness	3.82	3.61, 4.03	1.08	−0.42	1.29	6.00	.79
Neuroticism	4.64	4.38, 4.90	1.33	−0.62	1.00	7.00	.84
Openness	4.25	4.05, 4.45	1.04	0.01	1.67	6.83	.78
Agreeableness	4.61	4.41, 4.80	1.00	−0.29	1.83	6.67	.73
Self-Control	2.83	2.70, 2.95	0.66	0.02	1.08	4.31	.76
GPA	2.59	2.45, 2.74	0.77	−0.19	0.70	3.87	n/a
L2-GPA	2.54	2.37, 2.72	0.90	−0.30	0.57	3.92	n/a
Proficiency	502.83	474.75, 530.91	145.79	0.55	225.00	970.00	n/a

Regarding the reliability of the L2-Grit scale (research question 1), both consistency of interest (COI) and perseverance of effort (POE) of the L2-Grit scale (hereafter L2-COI and L2-POE) showed adequate levels of reliability in this study (L2-COI: α = .76, .66, L2-POE: α = .89, .86).

3.2 Construct validity of the L2-Grit scale

Research question 1 also asks about the validity of the L2-Grit scale. CFA was used to assess the goodness of fit of the two-factor model of L2 Grit. The results showed an adequate fit to the data: χ^2 (26) = 45.33, p = .011, x^2/df = 1.74, CFI = .96, TLI = .94, RMSEA = .08. Table 3 provides a summary of the standardized factor loadings observed in this study and the component loadings reported in the initial study. Here, one can first see that all factor loadings in the CFA model exceeded the threshold value of .30 (Harrington, 2009). In addition, the analyses performed in two different settings resulted in generally similar loading patterns. The similarity was observed for all L2-POE items (items 1 to 5 in Table 3) and two out of four L2-COI items (items 6 and 7). At the same time, notable differences were also found in the other two items: item 8 (.33, .67) and item 9 (.93, .53).

Stepwise regression analysis was performed to examine the degree of overlap between L2 Grit, and the Big Five traits and Self-Control. Conscientiousness emerged as the only significant predictor in this study (β = .22 [95% CI: .03, .41], SE = 0.07, t = 2.32), and the model explained 5% of the variance in L2 Grit (R^2 = .05, F (1, 104) = 5.38, p = .022, $1−\beta$ = .65). The normality of the residual distribution was checked using the K-S test (D = 0.08, p = .520). In contrast, in the initial study, a total of 10% of the variance in L2 Grit was explained by the Big Five traits (R^2 = .10, F = 4.30, p < .01), and the significant predictors were Extraversion (β = .21) and Emotional stability (β = .23).

3.3 Criterion-related and incremental validity of the L2-Grit scale

Partial correlation analysis was performed to evaluate how L2 Grit and other performance predictors (Grit, Conscientiousness, and Self-Control) are related to the performance measures (GPA, L2-GPA, and Proficiency). Table 4 summarizes the partial correlations between the personality traits and performance measures after controlling for the influence of gender and year at university. Here, Conscientiousness and L2 Grit had significant correlations with the performance measures.

Table 3. Standardized estimates for the two-factor L2 grit model and the component loadings reported in the initial study

		This study		Teimouri et al. (2022)	
Item		L2-POE	L2-COI	L2-POE	L2-COI
1	I will not allow anything to stop me from my progress in learning English.	.75		.92	
2	I am a diligent English language learner.	.79		.85	
3	Now that I have decided to learn English, nothing can prevent me from reaching this goal.	.71		.74	
4	When it comes to English, I am a hard-working learner.	.85		.70	
5	I put much time and effort into improving my English language weaknesses.	.79		.65	
6	I think I have lost my interest in learning English.		.84		.81
7	I have been obsessed with learning English in the past but later lost interest.		.58		.77
8	My interests in learning English change from year to year.		.33		.67
9	I am not as interested in learning English as I used to be.		.93		.53

Note. All estimates of this study were statistically significant at the .001 level.

In the initial study, Grit had weak correlations with GPA ($r = .14$), performance in the three L2 classes ($r = .13$ on average), and self-reported L2 proficiency ($r = .11$). Table 4 shows that the same trends were also seen in the current study ($r = .09–.15$). Another parallel can be seen in the correlation between L2 Grit and GPA ($r = .29$ in the current study and $r = .25$ in the initial study). Meanwhile, L2 Grit in the current study had a correlation of .40 with L2-GPA and .42 with Proficiency/the standardized test score. Both of these values are higher than the correlations previously obtained using the performance of the three L2 classes ($r = .27$ on average) and self-reported proficiency ($r = .31$) (Teimouri et al., 2022).

A series of hierarchical regression analyses were conducted to test the incremental prediction of the performance scores from Grit and Conscientiousness when L2 Grit was used as the first step regression model. Self-Control was excluded from the analyses owing to its very weak relationships with the performance measures (see Table 4). It should be again noted that the incremental validity of the L2-Grit scale was not assessed in the initial study.

Table 5 summarizes the results of hierarchical multiple regression analyses. The normality of the residual distributions was assessed using K-S tests (D = 0.06–0.09, $p = .323–.854$). The variance inflation factor values were all below 1.79. As shown in Table 5, L2 Grit was entered into the regression model in the first step in all cases. L2 Grit explained between 5% and 12% of the variance in the

Table 4. Partial correlations between personality trait scores and performance scores

Index	GPA	L2-GPA	Proficiency
Grit	.15 [−.06, .34]	.14 [−.05, .32]	.09 [−.11, .28]
L2 Grit	**.29** [.10, .47]	**.40** [.23, .56]	**.42** [.28, .55]
Conscientiousness	**.23** [.04, .40]	**.26** [.08, .42]	.14 [−.01, .29]
Self-Control	−.01 [−.21, .20]	−.01 [−.20, .20]	.02 [−.18, .20]

Note. Controlled for the influence of gender and year at university ($df = 102$). Bias-corrected and accelerated confidence intervals are presented (B = 5,000). Coefficients printed in bold are statistically significant at $p < 0.05$ (two-tailed), and $1 − \beta \geq .80$ when $r \geq 27$ or $\leq −.27$.

Table 5. Hierarchical regression analyses predicting performance scores with personality trait scores

DV	Model summary	IV	β	[95% CI]	SE	t	p
GPA	Step 1						
		L2 Grit	.23	[.04, .42]	0.10	2.37	.019
	$R^2 = .05$, F (1, 104) = 5.63, $p = .019$, $1-\beta = .65$						
	Step 2						
		L2 Grit	.21	[.01, .42]	0.10	2.08	.040
		Grit	−.09	[−.34, .14]	0.13	−0.67	.504
		C	.21	[−.01, .41]	0.12	1.70	.092
	$\triangle R^2 = .03$, F change = 1.53, $p = .222$						
L2-GPA	Step 1						
		L2 Grit	.34	[.18, .52]	0.09	3.76	<.001
	$R^2 = .12$, F (1, 104) = 14.12, $p < .001$, $1-\beta = .96$						
	Step 2						
		L2 Grit	.36	[.19, .57]	0.10	3.62	<.001
		Grit	−.17	[−.39, .02]	0.12	−1.43	.156
		C	.26	[.04, .45]	0.12	2.23	.028
	$\triangle R^2 = .04$, F Change = 2.51, $p = .087$						
Proficiency	Step 1						
		L2 Grit	.35	[.19, .50]	0.09	3.76	<.001
	$R^2 = .12$, F (1, 104) = 14.07, $p < .001$, $1-\beta = .96$						
	Step 2						
		L2 Grit	.35	[.19, .53]	0.10	3.56	<.001
		Grit	−.14	[−.38, .10]	0.12	−1.12	.264
		C	.21	[−.02, .40]	0.12	1.79	.078
	$\triangle R^2 = .03$, F Change = 1.60, $p = .207$						

Note. C = Conscientiousness, 95% CIs were estimated using bias-corrected and accelerated bootstrapping (B = 5,000).

performance measures (5% regarding GPA and 12% regarding L2-GPA and Proficiency). Adding Grit and Conscientiousness to the models increased the variance explained by 3–4%. None of the R^2 change values were significant, however. All these results indicate that L2 Grit was the most important predictor across all evaluated models.

4. Discussion

The results of Teimouri et al. (2022) provided support for the reliability and validity of the L2-Grit scale. The present conceptual replication assessed the generalizability of their findings using a different sample and different methodology. The methodological changes included:

1. The factorial validity of the L2-Grit scale was assessed using CFA.
2. The discriminant validity of the L2-Grit scale was tested using locally validated personality measures.
3. The behavior of L2 grit was assessed in relation to overall performance in L2 classes and standardized test score in addition to overall GPA.

4. The behavior of L2 grit was compared not only to that of grit but also to those of Conscientiousness and Self-Control.

The results obtained in this study broadly support the reliability and validity of the L2-Grit scale. The following will discuss the commonalities and differences between the results of this study and those of the initial study.

The first research question explored the reliability and validity of the two-factor structure of the L2-Grit scale. Reliability analyses in both studies showed that the internal consistency of L2 Grit and L2-POE was satisfactory ($\alpha > .70$). Furthermore, the results of the CFA showed that the two-factor structure provided an adequate fit to the present data. Meanwhile, two notable differences were found regarding the consistency sub-scale. The first is its reliability ($\alpha = .76$ in this replication and .66 in the initial study) and the second is the loadings of two items (i.e., items 8 and 9 in Table 3) (=.33 and .67, and .93 and .53). These differences suggest that more attention could be paid to understanding how learning contexts, including social and cultural settings, are relevant to the structure and internal consistency of the consistency sub-scale. Overall, however, the results of the current conceptual replication support the reliability and factorial validity of the L2-Grit scale.

Regarding the discriminant validity of the L2-Grit scale, both the present study and initial study found that the Big Five traits explain little variance in L2 Grit. The key difference was that Big Five Conscientiousness emerged as a significant predictor in this study ($\beta = .22$), while the initial study reported Extraversion and Emotional Stability as significant predictors ($\beta = .21$ and .23). On this point, the estimates gained in the current study are considered to be more robust in that the analysis was based on variables with adequate internal consistency. Also, the relationship found in the current study is more compatible with the results of meta-analysis, where grit, a similar but broader construct than L2 grit, shared a large amount of variance with Conscientiousness (Credé et al., 2017). It may therefore be more appropriate to consider L2 grit as a domain-specific variation of Conscientiousness. At the same time, however, the weakness of this study is that the present sample ($N = 106$) was not adequately powered to detect an effect size of $R^2 = .05$ (in this study, $1–\beta = .80$ when $R^2 \geq .07$ for a single-predictor model). For this reason, the present findings from the stepwise regression should be seen as tentative rather than definitive. With this limitation in mind, the results of the present replication provide additional support to the idea that the concept of L2 grit is different from other personality traits.

The second research question addressed the relationship between L2 grit and language achievements. All in all, the results of the present study provide strong support for the criterion-related and incremental validity of the L2-Grit scale. To begin with, the results of the partial correlation analyses were compatible with those of the initial study in that L2 Grit always showed stronger correlations with the performance measures than Grit.

The current conceptual replication also extended the findings of the initial study in three ways. First, L2 Grit had stronger associations with the three performance measures than Conscientiousness and Self-Control did. These results underline the merit of using the concept of L2 grit in SLA research. Discussion on the value of grit compared with Conscientiousness and self-regulation will continue (Credé et al., 2017; Ivcevic & Brackett, 2014), but individual differences in linguistic achievements can be most clearly explained by using the concept of language-domain-specific grit.

The second major implication is the relevance of L2 grit for long-term L2 learning. The initial study indicated that L2 grit was related to English majors' GPA and performance in the three target language classes. On this point, overall GPA reflects learners' long-term efforts in both L2 and non-L2 courses, and performance in several classes can be improved by narrowly focused efforts. The relationship observed between L2 Grit and L2-GPA, on the other hand, indicates that gritty learners are consistently more successful in language courses, as continuous efforts in multiple ESL classes are required for higher L2-GPA scores ($M = 2.90$ for the years of university-level L2 learning and $M = 24.43$ for the number of credits attempted).

Lastly, the use of a standardized L2 test enabled the current study to conclude that gritty language learners are more likely to be at a higher linguistic level in their L2. This finding is encouraging given

that the scores of standardized tests are often used for high-stakes decisions, including school admissions, class placements, and promotions. The results presented here suggest that language classes may contribute more to L2 success through the improvement of test scores, as teachers begin to pay close attention to the promotion of gritty behavior in addition to the development of L2 knowledge and skills. In the development of such interventions, the L2-Grit scale helps teachers examine how effective their designs are in terms of their learners' changes in language-domain-specific grit.

5. Limitations and future directions

Several limitations of the current study need to be acknowledged. The first is the sample size of this study, which limited its statistical power. The current study measured L2 proficiency by using a standardized test. Because the testing alone took two hours to complete, the final sample size was limited to 106, which is relatively small within relevant L2 literature (for a similar sample size, see Feng & Papi, 2020).

The second limitation concerns the correlational nature of the present data. Grit in SLA literature has typically been studied using data collected at a single time point. This study took the same approach and assessed the concurrent validity of L2 Grit, rather than discussing the issue of causality. In other words, the data presented here are correlational, thereby limiting conclusions as to whether L2 Grit causes the differences in linguistic achievement.

The last limitation relates to the measurement of L2 proficiency. The use of a standardized test has led to new insights on the relationship between L2 Grit and absolute levels of attainment. Meanwhile, it is also true that L2 proficiency in this study was estimated based on limited aspects of language skills (comprehension skills in particular). Therefore, there is still room to explore the relationship between L2 grit and language proficiency when, for instance, productive, interactive, and mediating skills are brought into the discussion.

Overall, the results of this study encourage the use of large sample sizes, longitudinal analyses, and various proficiency measures. These modifications will allow future replication studies to more fully describe the role of L2 grit in long-term language learning. A large sample study is challenging to conduct when employing time-consuming tests; this challenge, however, can be overcome by collecting data at multiple sites and paying careful attention to demographic differences between groups. In addition, longitudinal investigations with repeated measurements will help establish the cause-effect relationship between L2 grit and the development of various language skills.

6. Conclusion

This conceptual replication study re-examined the reliability and validity of the L2-Grit scale. Reliability and construct validity were assessed using reliability analysis, CFA, and stepwise regression analysis. The results were somewhat different from those of the initial study, but taken together they lend support to the internal consistency and construct validity of the L2-Grit scale. This study then examined the criterion-related validity employing two novel performance indexes (i.e., L2-GPA and objective L2 proficiency) in addition to a common GPA score, while simultaneously paying due attention to the incremental validity of L2 Grit compared with other personality traits. The results of the hierarchical regression analyses identified L2 Grit as the most important predictor across all evaluated models, and hence provided further evidence for the criterion-related and incremental validity of the L2-Grit scale.

Supplementary material. To view supplementary material for this article, please visit: https://doi.org/10.1017/S0261444822000544.

References

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, *25*(1), 60–83. doi:10.3102/10769986025001060

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. doi:10.1037/pspp0000102

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge.

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. doi:10.1037/0022-3514.92.6.1087

Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit–S). *Journal of Personality Assessment*, *91*(2), 166–174. doi:10.1080/00223890802634290

Ebadi, S., Weisi, H., & Khaksar, Z. (2018). Developing an Iranian ELT context-specific grit instrument. *Journal of Psycholinguistic Research*, *47*(4), 975–997. doi:10.1007/s10936-018-9571-x

Eskreis-Winkler, L., Duckworth, A. L., Shulman, E., & Beal, S. (2014). The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology*, *5*, 36. doi:10.3389/fpsyg.2014.00036

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi:10.3758/BF03193146.

Feng, L., & Papi, M. (2020). Persistence in language learning: The role of grit and future self-guides. *Learning and Individual Differences*, *81*, 101904. doi:10.1016/j.lindif.2020.101904

Ferguson, C. J. (2016). An effect size primer: A guide for clinicians and researchers. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 301–310). American Psychological Association. doi:10.1037/14805-020

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. Jr., (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. doi:10.1016/S0092-6566(03)00046-1

Harrington, D. (2009). *Confirmatory factor analysis*. Oxford University Press.

Ivcevic, Z., & Brackett, M. (2014). Predicting school success: Comparing conscientiousness, grit, and emotion regulation ability. *Journal of Research in Personality*, *52*, 29–36. doi:10.1016/j.jrp.2014.06.005

Jachimowicz, J. M., Wihler, A., Bailey, E. R., & Galinsky, A. D. (2018). Why grit requires perseverance and passion to positively predict performance. *Proceedings of the National Academy of Sciences*, *115*(40), 9980–9985. doi:10.1073/pnas.1803561115

Khajavy, G. H., MacIntyre, P. D., & Hariri, J. (2021). A closer look at grit and language mindset as predictors of foreign language achievement. *Studies in Second Language Acquisition*, *43*(2), 379–402. doi:10.1017/S0272263120000480

Namikawa, T., Tani, I., Wakita, T., Kumagai, R., Nakane, A., & Noguchi, H. (2012). Development of a short form of the Japanese big-five scale, and a test of its reliability and validity. *The Japanese Journal of Psychology*, *83*(2), 91–99. doi:10.4992/jjpsy.83.91

Ozaki, Y., Goto, T., Kobayashi, M., & Kutsuzawa, G. (2016). Reliability and validity of the Japanese translation of brief self-control scale (BSCS-J). *The Japanese Journal of Psychology*, *87*(2), 144–154. doi:10.4992/jjpsy.87.14222

Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

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. doi:10.1080/17501229.2020.1802468

Takehashi, H., Higuchi, O., Ozaki, Y., Watanabe, T., & Toyosawa, J. (2019). Reliability and validity of the Japanese version of the grit scale. *The Japanese Journal of Psychology*, *89*(6), 580–590. doi:10.4992/jjpsy.89.17220

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324. doi:10.1111/j.0022-3506.2004.00263.x

Teimouri, Y., Plonsky, L., & Tabandeh, F. (2022). L2 grit: Passion and perseverance for second-language learning. *Language Teaching Research*, *26*(5), 893–918. doi:10.1177/1362168820921895

Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2019). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, *22*(3), 516–536. doi:10.1017/S1366728918000421

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

Wei, Y., & Low, A. (2017). Monitoring score change patterns to support TOEIC® listening and reading test quality. *ETS Research Report Series*, *2017*(1), 1–20. doi:10.1002/ets2.12186

Hitoshi Mikami is an associate professor in the Department of English Language and Culture at Chubu University. His research interests include positive psychology, anxiety, self-regulation, and language assessment. His published work has appeared in journals such as *Reading in a Foreign Language* and *Frontiers in Psychology*.