JOURNAL FOR THE PSYCHOLOGY OF LANGUAGE LEARNING

IAPLL

Replication

Reexamining the "L2 Grit Scale" Construction Process: A Conceptual Replication of Teimouri et al. (2022)

Martiniano Etchart, *University of Wisconsin-Madison, United States*

https://orcid.org/0000-0002-7199-2102

metchart@wisc.edu

Paula Winke, *Michigan State University, United States; University of Innsbruck, Austria*

https://orcid.org/0000-0002-8169-650X

winke@msu.edu

ABSTRACT

Teimouri et al. (2022) created an influential second language (L2) grit scale that researchers have been using to investigate grit in second language acquisition (SLA) research. Based on Duckworth et al.'s (2007) domain-general grit scale, they drafted 20 L2 grit items which they piloted on 35 first language (L1)-Persian English learners. The authors then used the data to reduce the items from 20 to 12 to test on 191 new L1-Persian English learners. After analyses, they finally reduced the items to 9 comprising two components: *perseverance of effort* and *consistency of interest*. In our study, we investigated whether we could reproduce the scale's item-reduction from 20 to 9. We had 580 L1-English learners of Spanish take the original 20-item, L2 grit survey. Different from Teimouri and his colleagues, our analyses did not lead us to the same 9 items; rather, 17 were retained with three factors. Through post-hoc tests, we explored whether Teimouri et al.'s final 9-item, L2 grit model with two constructs would fit our new data using structural equation modeling. While it fit, parameter-estimates indicated the two constructs may be mirror opposites, which may problematize score interpretation and construct definition, leading us to reflect on the three-factor solution as perhaps a better fit for our data. Grit may certainly be important for L2 learning, as researchers are finding by using the 9-item scale. However, researchers should be aware that the field may not have a definitive way to measure L2 grit. Thus, we claim that researchers should continue to investigate how to best measure L2 grit, explore alternatives, and publish their data with their L2 grit research, so that the scales can be better examined and scores from them better interpreted.

Keywords: L2 grit, perseverance, interest, scale construction, replication, reproducibility

Journal for the Psychology of Language Learning
ISSN: 2642-7001
(2024) Volume 6, Issue 2. e6216324.
https://doi.org/10.52598/jpll/6/2/1

INTRODUCTION

Grit has been viewed as "one of the most important traits that any individual should possess to be successful in the 21st century" (U.S. Department of Education, 2013, as cited in Changlek & Palanukulwong, 2015, p. 26). Within the field of second language acquisition (SLA), however, the question of whether *grit* is a factor that is related to or promotes successful second language (henceforth, L2) learning has only recently been asked. Duckworth et al. (2007) defined grit as "perseverance and passion for long-term goals" (p. 1087). Duckworth and her colleagues added that grit "entails working strenuously toward challenges, maintaining effort and interest over years despite failure, adversity, and plateaus in progress" (2007, pp. 1087–1088). Thus, building grit and setting concrete, ultimate goals could determine whether students will successfully achieve what they would like to achieve and, consequently, whether they will thrive while doing it.

In the late 2010s, after Duckworth (2016) published her popular New York Times bestselling book on grit, researchers in applied linguistics began focusing on explorations on grit and in determining its role in L2 development. Two early studies were conducted with children learning English as an L2: First, researchers Banse and Palacios (2018) used an 8-item 5-point Likert scale taken from the Short Grit Scale (Duckworth & Quinn, 2009) to measure 3,272 elementary Latino students' grit and its relationship with English language arts (ELA). They found a significant three-way interaction between grit, English language learner (ELL) status (a student defined as an English learner or not in the public school), and classroom characteristics (care and control), indicating that student perceptions of care and control were particularly salient for the students' grit. The authors also highlighted the predictive validity of grit on students' ELA achievement.

Secondly, Lee (2020) employed Duckworth et al.'s (2007) 10-item grit scale to gauge L2-English-learning South Korean students' (n = 137 middle school, n = 323 high school, and n = 187 university students) perseverance of effort and consistency of interest. Lee investigated the relationship between grit, classroom enjoyment, and students' willingness to communicate in the target language (L2 WTC). Lee found that grit and classroom enjoyment were significant predictors of participants' L2 WTC behavior in the English-as-a-foreign-language (EFL) classes. However, whereas perseverance of effort was a strong predictor of all students' L2 WTC, consistency of interest did not predict L2 WTC in all groups. Conclusively, Lee emphasized the importance of applying positive psychology constructs to expand SLA theory, a position promoted eloquently by MacIntyre et al. (2019) as well.

The early studies into L2 grit and its effects on learning were promising in demonstrating that L2 grit may be meaningful in SLA. However, the variation in L2 grit measurement instruments across the studies was a concern. A question arose: Should L2 researchers use the original, domain-general grit scale developed by Duckworth et al. (2007), or the short form of that same scale (Duckworth & Quinn, 2009)? A third option became possible based on the work of Teimouri et al. (2022): These pioneering researchers created an SLA field-specific, language-domain measurement scale of L2 grit based on Duckworth et al.'s (2007) domain-general grit scale. Their 9-item scale was first published as an appendix within their article's advanced online publication in 2020. The instrument was thus subsequently adopted and adapted by other SLA researcher groups who have used it to demonstrate, for example, (1) how L2 grit relates to other individual difference variables that are conceptually similar to grit and also important for learning (Sudina & Plonsky, 2021a); (2) how teachers' behavior can improve (Derakhshan et al., 2023) or (3) predict (Wu et al., 2023) learners' L2 grit; (4) how L2 grit is related to L2 achievement and anxiety (Sudina & Plonsky, 2021b); (5) how fostering grit can help L2 learners better regulate their feelings of anxiety (Alazemi et al., 2023); or (6) what may be the most important predictors of L2 grit in an online language learning setting (Paradowski & Jelińska, 2023).

Most recently, MacIntyre and Khajavy (2021) published a special issue on L2 grit. They wrote in their introduction that "as grit attracts the attention of SLA researchers, there is a need for deeper understanding of the construct, its potential and limitations" (p. 2). MacIntyre and Khajavy additionally stressed that since Duckworth's grit scales have been published (e.g., Duckworth et al., 2007; Duckworth & Quinn, 2009), there have been healthy and necessary questions about the scales and their measurement properties. Within the same issue, Oxford and Khajavy (2021) outlined how L2 grit measurement scales tend to have problems, such as measurement constructs that do not appear to align well with the theoretical constructs, the inclusion of

questionably-worded scale items, and the use of time metaphors in the items that may not relate to the test takers' experiences of time within their language learning trajectories. MacIntyre and Khajavy wrote that SLA researchers should follow advice by Teimouri et al. (2021) and "avoid the temptation to uncritically use any of the existing grit measures" (p. 4). They further wrote that "studies of grit in SLA need to consider revising and developing a collection of grit measures that reflect the main aspects of the grit construct" (MacIntyre & Khajavy, 2021, p. 5).

In this paper, we explore the developmental procedures related to the first domain-specific measure of L2 grit (see Appendix 1 for details on the scale developed by Teimouri et al., 2022) so that we can better understand it as an existing L2 grit measure, especially as it gains prominence and uses within the SLA field. More specifically, we try to reproduce Teimouri et al.'s (2022) study's data reduction results. Since Teimouri et al. used a series of data analyses to move the scale from 20-items to 9-items, we wanted to identify the stability of the 9-item selection from the 20 original ones. In the process, we further aimed to investigate the two main constructs of L2 grit, perseverance of effort and consistency of interest, which the 9 items represent.

LITERATURE REVIEW

The First L2 Grit Scale: Teimouri et al.'s (2022) 9-item L2 Grit Scale for SLA Research

To define a domain-specific measure of L2 grit for research within the field of applied linguistics, Teimouri et al. (2022) developed and provided evidence regarding the validity of the first language-domain-specific grit scale that helped measure students' levels of grit in language learning contexts. Teimouri et al. (2022) specifically focused on (a) students' consistency of interest by measuring students' unwavering interest during the L2 learning process; and (b) their perseverance of effort by measuring how persistent learners were in applying effort to achieve their goals regarding their (new) language being learned.

Through the implementation of this new L2 grit scale, Teimouri et al. (2022) demonstrated that the 9-item language-specific grit scale predicted motivational variables and exhibited stronger positive correlations to students' language learning behaviors and achievements than the domain-general grit scale. Furthermore, Teimouri et al. explained that the use of such "language-domain-specific measures of personality traits [in future research] – especially, those of immediate relevance to L2 learning – will provide more accurate findings and pedagogical implications" (p. 18). In addition, the authors claimed that the language-domain-specific measurement of grit will allow for clearer and more meaningful data for SLA research.

Subsequently, as we alluded to in this paper's introduction, the newly developed L2 grit scale has been adopted by multiple research groups to investigate various factors and processes in SLA. In Table 1, we summarize the main components of six studies that have used the 9-item grit scale. Considering that the 9-item L2 grit scale is a recently developed instrument to measure language learners' levels of grit towards L2 learning, and because interest among SLA researchers in using the 9-item L2 grit scale with or without adaptations appears to be increasing, we sought to investigate whether Teimouri et al.'s (2022) scale construction results could be replicated. Particularly, we focused on conducting an extension and reproduction of the scale-construction results from the original 2022 study by Teimouri and colleagues, by engaging a different sample of L2 learners.

We were interested in this research because one of the limitations mentioned by Teimouri et al. (2022) was that the participants in their study were students majoring in English who may have had higher levels of L2 grit because "[their] future careers are dependent on their English language skills" (2022, p. 17). In fact, in one of the implementations of Teimouri's L2 grit scale, Sudina and Plonsky (2021a) hypothesized that the L2 was one of the reasons why consistency of interest (rather than perseverance of effort, as claimed by Teimouri) was a superior predictor of L2 achievement. Sudina and Plonsky explained,

> students' sustained interest—rather than perseverance—is a driving force for learning French and Spanish at a U.S. college due to their genuine interest in the culture, history, and literature of French- and Spanish-speaking countries. This is not necessarily the case with EFL, the learning of which is often considered crucial for advancing one's education and career. (Sudina & Plonsky, 2021a, p. 844)

In order to address the question of how generalizable the structure of the L2 grit scale is, our sample included 580 Spanish-language learners (n = 36 majoring in Spanish) in the United States, a different sample and a different language-learning context than in Teimouri et al.'s (2022) study. We had three additional reasons for replicating Teimouri et al.'s work. First, we wanted to see if we could replicate the data reduction described in Teimouri et al. because we had some questions about the data reduction steps. In Teimouri et al.'s published paper, the final questionnaire had 9 items, which were reduced from their original questionnaire with 20 items in two steps: First, the authors piloted the 20 items on 35 first-language (L1)-Persian learners of English and used classical test theory and principal component analysis (PCA) to reduce the 20 items to 12.

Table 1. *Published Studies That Have Used Teimouri et al.'s (2022) 9-Item L2 Grit Scale to Investigate Factors or Processes in SLA*

Study	SLA Research Aim(s)	Participants	Main Finding(s)
Sudina and Plonsky (2021a)	To investigate whether L2 grit differs from conceptually related constructs and how all of them relate to L2 achievement	360 university Spanish (n = 258) and French (n = 102) learners in the United States	L2 grit, L2 buoyancy, intended effort, and conscientiousness (industriousness)
Derakhshan et al. (2023)	To measure the effects of student-perceived teacher affective behavior (namely, teacher support, teacher enthusiasm, and teacher appreciation) on students' L2 grit	285 L1-Turkish EFL university students	Strong correlations among all L2 teacher variables, especially teacher enthusiasm, with learners' L2 grit (both perseverance of effort and consistency of interest).
Wu et al. (2023)	To investigate whether perceived teacher affective support and perceived teacher enjoyment were good predictors of academic burnout with L2 grit as a mediator	1,294 Chinese college students learning English	L2 grit acted as a partial mediator between teacher affective support and burnout, but not perceived teacher enjoyment: in sum, L2 grit was boosted by teacher affective support and helped prevent academic burnout.
Sudina and Plonsky (2021b)	To L2 grit's relationship with language learning anxiety, L2/L3 achievement, and self-rated proficiency	153 college students studying two languages in Russia	L2 grit's perseverance of effort and language learning anxiety predicted Russian students' achievement and self-rated L2 proficiency.
Alazemi et al. (2023)	To investigate whether L2 grit (in addition to academic emotion regulation, resilience, and self-assessment) is related to learners' levels of test anxiety	417 EFL university students from Kuwait	Learners with higher L2 grit were better at modulating their test anxiety and had higher levels of self-regulation and confidence.
Paradowski and Jelińska (2023)	To identify whether language mindsets, curiosity, autonomy, and readiness for online learning would affect L2 grit in virtual education settings	615 adult learners from around the world taking languages (n = 33) hybrid or fully online	Readiness for online learning (composed by both learning motivation and self-directed learning) was the main determinant of L2 grit in virtual as well as hybrid environments. Curiosity was also associated with L2 grit, while mindsets were not as much.

Note. L1 = first language, L2 = second language, L3 = third language; EFL = English as a foreign language.

Second, the authors had 191 L1-Persian learners of English take the 12-item L2 grit scale. The authors followed the same procedures to reduce the number of items to 9. In the paper, in a paragraph on the 12-item, L2 grit scale, the authors reported that the scale is on the Instruments and Data for Research in Language Studies (IRIS) digital repository and is also in their (Teimouri et al., 2022) Appendix 1. At those locations, we found the 9-item grit scale, but not the 12-item scale. We became curious as to how the data reduction unfolded from 20, to 12, to 9. We sought to obtain

the study's original 20-item scale. Through personal communication with the first author of Teimouri et al.'s study, we obtained the original 20-item scale.

Second, we were curious about how Teimouri and his colleagues (2022) removed 3 items (from 12 to 9) after running item-total correlations analysis. In the paper, they explained that these three items fell below the minimum criteria of .40, but it was unclear in the original study if the minimum of .40 was an absolute value (that is, if removal was due to the item being between –.40 and .40), or if all items below a positive value of .40 were removed. We assumed that the authors meant that the absolute value was used as the criteria for removal, but we wanted to check the data to be sure. Because we did not have access to the data, it was necessary to collect new data to check such assumptions. Or, at the very least, since we are interested in using the 9-item L2 grit scale ourselves for new research, we thought as a first step, we would first scrutinize the 9-item scale's construction, to ensure that our future usage of it would produce maximally interpretable scores within our study's context, that is, learners of L2 Spanish.

Third and lastly, we wanted to scrutinize this: Teimouri et al. (2022) did not include the results of all of the loadings of the 9 items onto their respective components (i.e., *effort* and *interest*) when presenting the outcome of the PCA in Table 3 (p. 13). Thus, half of the item loadings were suppressed in the results, which is common in prior research, but not transparent enough today, we believe, especially when the data are the basis of a proposed standardized and shareable scale for any L2 researcher, and when the raw data are also not available to the public. We found data suppression to be an important aspect to consider for validity and reliability purposes. Suppression within the output makes the data easier to read, but such suppression also makes it difficult to fully comprehend the overall picture. When items load above the loading-threshold across two or more components or factors, researchers have the choice to (a) eliminate the item (because it does not reliably load on a single factor), or (b) accept one of the two high loadings based on loading strength. In the latter case, researchers can either (a) choose the higher of the two loadings or, for purposes of theory alignment, (b) choose the one that makes the most theoretical sense. We wanted to *see* the values in the blanks where loadings were not reported, so that we could scrutinize the loadings and any choices the researchers may have made in cases of double or multiple loadings. In other words, we questioned whether we fully understood the L2 grit instrument given the lack of complete, data-backed evidence for its validity arguments. We wondered whether this is an instrument that is ready to be used by us or other researchers in future studies, or if the scale itself could use further research, refinement, or use qualifications, as others have suggested.

The Need to Replicate Aspects of Teimouri et al.'s (2022) Work

We are not the first to replicate scale-composition aspects of Teimouri et al.'s (2022) work, which stresses the importance and influence of their contribution and the scale they created. At least three studies have been published that re-investigated the constructs of the L2 grit scale (Elahi-Shirvan et al., 2021; Mikami, 2023; Wei et al., 2020).

First, Wei et al. (2020) had 462 diverse Chinese learners of English, ages 18 to 52, take Teimouri et al.'s (2022) 9-item L2 grit scale to investigate its underlying construct on a sample different from the original study. In the Chinese EFL context, the scale proved as reliable (Cronbach's α of .80) as in Teimouri et al.'s study (also Cronbach's α of .80). Like Teimouri et al., Wei et al. used PCA to calculate the loadings of the 9 items and to identify their underlying structure. The 9 items loaded across two components in the same way as in Teimouri et al., and each item had a loading on their main component at .76 or higher. Like in Teimouri et al., loadings lower than .4 were suppressed, and data from the study are not publicly available for review.

Second, Elahi-Shirvan and colleagues (2021) conducted a longitudinal confirmatory factor analysis-curve of factors model (LCFA-CFM) to measure 437 Iranian EFL learners' L2 grit over time. They used a CFA model of Teimouri et al.'s (2022) L2 grit scale and checked the scale's internal consistency through Cronbach's α (= .87) as well as its composite reliability by applying McDonald's ω (= .88). Responding to a call for analytical approaches that captured the complexity and dynamism of individual differences in SLA research (Hiver & Al-Hoorie, 2019), and to address "the shortcomings of the validation of the existing [L2 grit] scale and the need to reconsider its construct validity" (Elahi-Shirvan et al., 2021, p. 1470), Elahi-Shirvan and colleagues' main goal was to attend to the temporal and developmental nature of L2 learners' grit. The authors concluded that the L2-grit factor and its two indicators (perseverance of effort

and consistency of interest), did not vary across the four points in time considered in the study. Additionally, they found that, as a continuous variable ranging from low to high, L2 grit was affected by other variables such as teachers' support, especially in less gritty students. Elahi-Shirvan et al. concluded that due to the context-specific nature of grit, further components could be added to the modeling of the construct in future studies.

Finally, Mikami (2023) published a conceptual replication of Teimouri et al.'s (2022) validation of their 9-item L2 grit scale. Mikami had 106 English majors at a Japanese university take the 9-item L2 grit scale. Mikami used confirmatory factor analysis (CFA) to analyze the results, which may be a more appropriate analysis of item loadings and their underlying structure than PCA is when the theory of how the data should fit together has already been proposed, as it had been by Teimouri et al.'s work. Mikami's CFA showed that the two-factor model of L2 grit with 9 items fit adequately to the data. Model fit is an important construct, especially in terms of standardized testing: Testing for model fit is somewhat like testing the design of a new car. It is like asking "Can this car be driven with this fuel?" The car's system of parts (constructs) and engineering (connectors or parameters) is the model.

The data is the fuel being put into the car to run the test drive. Adequate fit means the car started and it ran (it indeed worked), but the level of fit (adequate) means it was a somewhat clunky ride, which may be improved with either a different (improved) model, or with a different data set (fuel). In Mikami, the factor loadings were suppressed and not reported if they were lower than .30. Mikami found the 9 items loaded onto the two-factor structure in the same way as they did in Teimouri et al. and in Wei et al. (2020).

The aforementioned research results are unable to suggest whether a different combination of the original 20 items might prove an equal, different, or even better measure of L2 grit. Moreover, the unavailability of the data across the studies makes it difficult to scrutinize the scale at a level that may be required if the scale is to be reused often, across contexts, and with learners at varying levels of proficiency. With a standardized scale, scale-users may be entitled to triangulated evidence-based support of the use of the scale beyond its original application. Each study conducted to validate the scale, or to research other models of scale-development or scale-adaptation, should, over time, add to the field's better understanding of L2 grit.

We designed this replication of Teimouri et al.'s (2022) research to be a conceptual replication of their L2 grit scale-construction process. Conducting a constructive (or conceptual) replication:

> means beginning with a similar problem statement as the original study but creating a new means or design to verify the original findings. (…). Thus, for example, different, but related, measures (…) will help add to the body of knowledge obtained in the original report by validating the outcomes using two different techniques. Successful constructive replications provide stronger support for the original theory or hypothesis since evidence is provided that the outcomes are not limited to one particular methodology used. (Language Teaching Review Panel, 2008, p. 14)

We decided to begin at the same point Teimouri et al. (2022) did, with a 20-item L2 grit scale, and we sought to verify the construction of the 9-item scale. Table 2 is a detailed description of the research areas we conceptually replicated and how our data reduction plan differed from Teimouri et al.'s. Table 2 notably includes this study's research questions, positioned next to Teimouri et al. for contrastive purposes.

METHOD

Participants

A total of 580 Spanish language learners studying Spanish at a public university in the United States participated in this study. We originally collected data from a total of 625 Spanish language learners in fall 2020. After data collection, 45 students were removed from the study before analysis for not having answered all of the 20 fall 2020 L2 grit questions in our survey, leaving a sample size of 580. The sample of 580 students consisted of 403 students who self-identified as females, 160 who identified themselves as male, 12 who identified as something other, such as neither, non-binary male or female, gender non-conforming, or non-preferential pronouns, and 5 who left the question on gender blank. The sample's ages ranged from 18 to 47 years old at the time of

testing. The starting age of acquisition of the 580 students ranged from 0 to 27 years (M = 12.5; SD = 4.0).

Their language proficiencies ranged from beginner to upper-intermediate, which we interpret as most probably representing novice-low to intermediate-low or intermediate-mid on the American Council on the Teaching of Foreign Languages (2024) proficiency scale based on their placement into first-year, first semester 101 (n = 89), first-year, second semester 102 (n = 102), first-year, fast track 150 (n = 20), second-year, first semester 201 (n = 98), second-year, second semester 202 (n = 90), and third-year, first semester 310 (n = 175) classes (six learners did not record their placement). The learners' full demographics, but with no personally identifying information, and their scores are in this study's data file that we published online in the Open Science Framework (Etchart & Winke, 2023).

Materials

To collect data for this study, we used Teimouri et al.'s (2022) original 20-item L2 grit survey, two measurements of academic achievement, and information on the language learners' prior-to-college L2-Spanish learning backgrounds. We describe these data collection measures next.

Language-Domain-Specific L2 Grit Survey

We used the language-domain-specific grit scale developed by Teimouri et al. (2022) to measure L2 Spanish learners' grit and examine how it manifests itself in another language-learning setting. We used Teimouri et al.'s original survey comprising 10 items to measure the levels of *persistence of effort* (e.g., "When it comes to Spanish, I am a hard-working learner.") and 10 items measuring *consistency of interest* (e.g., "I try to keep myself motivated while learning Spanish."). We slightly adapted the wording of some statements by removing terms, such as think, always, or often, to avoid ambiguity, a common scale-construction recommendation, as discussed by Krosnick and Presser (2010).

Differently than Teimouri et al.'s (2022) L2 grit scale, we changed the number of Likert-scale points from 5 points to 10, with the new scale using the same endpoint descriptors: 1 was equated with *definitely not like me*, and 10 with *definitely like me*. We increased the scale to 10 based on scale-construction recommendations from within the field of educational measurement (Preston & Colman, 2000): Our move to 10 points was meant to increase measurement variance and scale reliability, which we wanted to do to help us uncover how variance in one construct (L2 grit) is related to variance in another construct (L2 achievement).

We expected the changes along these parameters to be minorly scale-improving. Moreover, the larger Likert-scale can be collapsed down to a smaller Likert-scale for meta-analytic or comparative-study purposes. We believed the scale would not affect the final reduction to 9 that we expected.

Language Background and Achievement Measures

To better examine the relationship between L2 grit and students' success in learning the Spanish language, we collected language background and achievement data from learners: (1) students' age of onset (AO), (2) enrollment in Spanish courses at university level, (3) expected grade in their current Spanish course, and (4) their overall, current (at the time of the questionnaire) college-level, grade-point average (GPA). Participants were also asked about whether they were majoring or planned to major in Spanish. We additionally collected information on whether learners were heritage language learners of Spanish (n = 33 in the overall 625 set of participants; n = 32 in the 580 subset of participants), but because there were few of them, we did not separate them out in the analyses. The heritage language learners are marked, however, in the full dataset, which can be found at Etchart and Winke (2023). All variables were computed based on different scales respectively (e.g., *0.0* through *4.0* for GPA). However, students' responses to the question on their (Spanish) major was assessed dichotomously, as *yes* or *no*, and later converted to a 0 (*no*) and 1 (*yes*) dichotomous code for analysis purposes.

Table 2. *Comparison of the Components of Teimouri et al. (2022) and This Study*

Components	Teimouri et al. (2022)	This Study
Participants	English as a foreign language learners (*n* = 191) (with an additional 35 learners who formed a pilot group, used to reduce the scale from 20 to 12)	Spanish as a foreign language and heritage language learners (*n* = 580)
Starting Instrument	20-item L2 grit scale, through a 5-point Likert scale (1 = *Not at all like me*, 5 = *Very much like me*)	20-item L2 grit scale, through a 10-point Likert scale (1 = *Definitely not like me*, 10 = *Definitely like me*)
Item examples, with minor changes	I think I have lost my interest in learning English (Interest).	I have lost my interest in learning Spanish (Interest).
Other types of measurements	Domain-general grit, teacher perception, intended effort, L2 WTC, attention, mindset type, L2 anxiety, L2 joy, L2 achievement	L2 achievement, learning context, form of instruction
RQs	RQ1: How valid and reliable is the newly developed L2 grit scale in measuring learners' perseverance and passion for L2 learning and use? RQ2: How is L2 grit related to L2 learners' motivation and emotions? RQ3: How is L2 grit related to language achievement?	RQ1: When we follow a data-reduction plan similar to that used by Teimouri et al., but with new data, do we arrive at the same 9 L2 grit items and two constructs? RQ2: How does Teimouri et al.'s (2022) model of L2 grit fit with data from Spanish language learners?
Analyses	Classical test theory and PCA	EFA and SEM

Note. RQ = research question, L2 = second language, WTC = willingness to communicate, PCA = principal component analyses, EFA = exploratory factor analysis, SEM = structural equation modeling.

Procedure

Participants were recruited with volunteer help from their own Spanish instructors, who were contacted by the first author of the study (who was at the time a Spanish instructor himself at the university). After giving their consent to participate, participants took a Qualtrics survey, in which they completed both the 20-item L2 grit scale and the achievement and background questionnaire. Those participants who reached the end of the Qualtrics survey received one extra credit in their Spanish language class as compensation for their participation in the study. At the end of the survey, they were directed via an external link to a sign-up sheet for the extra credit. The sign-up sheet consisted of a Google form that was unconnected to the survey data to ensure that no one, not even the researchers, were able to link the students' names to their responses on the survey.

Analysis

As our first analysis step, we calculated descriptive statistics for the 20-item L2 grit scale using the 2020 survey data from the 580 learners, and then we ran an EFA on the 20 items from that group to understand whether one of the main conclusions from Teimouri et al. (2022) would hold with a new set of data. In other words, we ran an EFA to uncover if we would find that L2 grit is a 2-factor latent trait with 5 items related to one factor that can be called *persistence of effort*, and 4 items related to a second factor that can be called *consistency of interest*. We ran our data with EFA rather than PCA because we wanted to investigate L2 grit as a multi-component latent trait, as defined by Duckworth et al. (2007), and PCA is not recommended for latent-trait research (Widamen, 1993).

As described by Garnaat and Norton (2010), "[w]hile PCA is commonly used in the literature, the assumptions made by this technique are generally not suited for use with psychological questionnaire data" (p. 723), and this is because PCA does not assume measurement error, which psychological questionnaires are assumed to have. Factor

analysis is more appropriate because it assumes measurement error and can identify groupings among items and reveal if items are superfluous to the construct. We believe EFA may be a more precise method for examining the structure of data caused by L2 grit, a latent trait, than PCA is, although, with larger datasets or large variable sets, EFA and PCA results are extremely comparable (see Widamen, 1993).

To help us identify the number of factors within the EFA, we investigated which factors had Eigenvalues greater than one, investigated factor alignment on a scree plot, and we additionally calculated mean Eigenvalues from randomly generated correlation matrices based on the size of the data set (580 participants and 20 variables) using an R software-based parallel analysis engine (Patil et al., 2017). These three methods combined allowed us to triangulate information for a better selection on the number of factors.

Secondly, as a post-hoc exploration, we directly addressed the generalizability of Teimouri et al.'s (2022) proposed theoretical model of L2 grit by using structural equation modeling (SEM) to explicitly test their 9-item L2 grit model on our samples of 580 learners. Within the structural model, we included a regression analysis—to match what Teimouri et al. did—to model the effects of L2 grit on achievement. By fitting our data onto the model, we were able to understand if the model of L2 grit proved stable with a new data set. Third and finally, we calculated correlations between the 580 learners' 9-item, L2-grit-scale factor scores and the learners' backgrounds and learning characteristics to explore further the potential relationships among L2 grit and other variables.

RESULTS

Replicating Steps Within Teimouri et al. (2022) to Test the Composition of the L2 Grit Scale

As we indicated above, our first step was to calculate descriptive statistics and reliability analyses for the 20-item L2 grit scale using the 2020 survey data from the 580 L1 English learners of Spanish. The results are displayed in Table 3. We calculated the reliability coefficient of L2 grit overall (.929), which is a calculation of Cronbach's α coefficient based on all 20 items together, because Teimouri et al. (2022) did the same, but we would like to point out that if the scale with 20 items has two factors, calculating an overall, single reliability for the test may be theoretically problematic. A better estimate of the overall scale reliability may be an average of the two factors' (item subsets') reliability estimates, which in this case would be .876.

Second, we ran an EFA on the 20 items we gave to the 580 Spanish language learners. The factor Eigenvalues and the total variance explained by the factors are in Table 4. The scree plot is in Figure 1. The item loadings on the three factors with Eigenvalues greater than 1 are in Table 5. In Table 5, we bolded item loadings that were greater than absolute .4 and we entered the results of a parallel analysis we computed using the total number of participants and the total number of variables.

We used a combination of the Eigenvalue criteria (checked for Eigenvalues greater than one), a visual inspection of the scree plot, and the information from the parallel analysis to determine the number of factors that should be retained. Notably different from Teimouri et al.'s (2022) final questionnaire is that the 20 items, when given to our Spanish language learners, did not result in two factors. Rather, three factors presented themselves: three factors' Eigenvalues were greater than 1; the parallel analysis did not rule out three factors; and the scree plot demonstrated three factors (a leveling off of the factors occurred after factor 3). Three of Teimouri et al.'s original five *perseverance of effort* construct items loaded on Factor 1 in this study: The other two loaded on Factor 2 in this study. All four of Teimouri et al.'s *consistency of interest* construct items loaded on Factor 3 in this study. If we were to recommend an L2 grit survey based on our EFA, we would propose researchers use the 17 items across the three factors that load higher than absolute .4. We labeled the three factors in this study "Effort 1: Internal and personal" (which could also be called *Positively worded perseverance of effort)*, "Effort 2: Overcoming external obstacles" (which could also be called *Negatively worded perseverance of effort*), and "Waning interest" (which could also be called *A lack of consistency of interest*). We labeled these three factors based on the content of the items within the factors, on prior naming conventions, and

identified constructs from Teimouri et al. and Duckworth et al. (2007).

A researcher wanting to ensure more clarity within and across the three constructs could choose to not use items with factor loadings lower than absolute .5 (that is, between –.5 and .5) (Backhaus et al., 2021, p. 417), which would reduce the number of items by three, eliminating two from *perseverance of effort* and one from *a lack of consistency of interest.* Another option could be to not use the items from *a lack of consistency interest*, as a matter of choice, if one can justify concentrating on the positive psychology aspects of L2 grit, thus bypassing the use of primarily negatively worded items, as reviewed as a consideration in measuring L2 grit by Oxford and Khajavy (2021). Another option, as employed by Mikami (2020), would be to keep items that load greater than absolute .3, which would result in all 20 items being retained.

Retaining all 20 items along Mikami's (2020) grounds could be justifiable because statistically, a scale with 20 items, as compared to one with 9, would be more reliable. These choices are matters of preference, theorizing, and the inspection of measurement properties relative to the data at hand. As mentioned by Tabachnick & Fidell (2001),

> most researchers use some guideline for a lower limit on item factor loadings and cross-loadings to determine whether to retain or delete items, but the criteria for determining the magnitude of loadings and cross-loadings have been described as a matter of researcher preference. (as cited in Worthington and Whittaker, 2006, p. 823)

In any case, the choices could be supported by statistical and qualitative evidence from research data. For data comparison across studies, the most important aspect may be including the questions and the data with the publication, so that if various studies do use different questions, those that are constant across studies could be used as anchors in meta-analyses.

SEM of Teimouri et al.'s (2022) Model of L2 Grit and how it Relates to Learning

The primary goal with our research was to replicate, as best as possible, the findings of Teimouri et al. (2022), who used a combination of PCA and regression analyses to (a) define the construct of L2 grit and to (b) understand the impact of L2 grit on language achievement measures. In essence, through their earlier (Teimouri et al., 2020) exploratory work, Teimouri et al. developed a hypothesized model of the relationships among two components of L2 grit, which they called (1) *perseverance of effort* and (2) *consistency of interest*, and measures of L2 achievement. We tried to arrive at their 9 items through EFA but were unable to do so. Does this mean the 9-item scale with two factors is uninformative? In exploration, we decided to diagram a SEM (see Figure 2) based on Teimouri et al.'s hypothesized model with the 9 items at which they arrived by the end of their research.

We designed the model specifically, so that we could fit our Spanish language learners' data to the model to see if the theoretical model would hold and be informative with our learners' data. We fit the model using maximum likelihood estimation through the software called Analysis of Moment Structures (AMOS 27) with our 580 learners' data. Next, we describe the model components so readers new to SEM can understand the model's structure.

Factor 1 (Perseverance of Effort) --> 5 Items on Effort

The theory in this model is that people have a characteristic (or embody a latent trait or factor) called *perseverance of effort*, and this personal characteristic can be adequately measured by the items numbered 1 through 5 in the model. The underlying conception is that one's amount of *perseverance of effort* will dictate or drive how people respond to the five items. In our research, learners responded to each item using a 10-point Likert scale, with 1 being *definitely not like me*, and 10 being *definitely like me*. Overall, high responses on the five items' Likert-scale (answers closer to 10 on the 1 to 10 scale) will mean that the person has a high *perseverance of effort,* which is one of two elements (or factors) of L2 grit.

Table 3. *Descriptives for the 20-item L2 Grit Scale (N = 580)*

Variable	Item	*M*	*SD*	Min	Max	Scale reliability (α)	95% CI lower	95% CI upper
L2 grit overall	(20 items' mean)					0.929*		
Factor 1: Persistence of effort	(10 items' mean)					0.881		
	1	7.05	1.99	1.00	10.00		6.89	7.22
	2	7.30	1.89	1.00	10.00		7.14	7.45
	3	7.04	2.38	1.00	10.00		6.85	7.24
	4	6.79	2.64	1.00	10.00		6.58	7.01
	5	5.02	3.17	1.00	10.00		4.76	5.27
	6	6.50	1.94	1.00	10.00		6.34	6.65
	7	7.49	1.85	1.00	10.00		7.34	7.64
	8	5.25	2.63	1.00	10.00		5.03	5.46
	9	7.02	2.35	1.00	10.00		6.83	7.21
	10	6.82	2.38	1.00	10.00		6.63	7.02
Factor 2: Consistency of Interest	(10 items' mean)					0.87		
	11	5.21	2.72	1.00	10.00		4.99	5.43
	12	3.95	2.60	1.00	10.00		3.74	4.16
	13	6.13	2.37	1.00	10.00		5.94	6.32
	14	5.92	2.49	1.00	10.00		5.72	6.12
	15	3.32	2.50	1.00	10.00		3.12	3.53
	16	4.11	2.79	1.00	10.00		3.88	4.34
	17	5.29	2.33	1.00	10.00		5.10	5.48
	18	7.50	1.89	1.00	10.00		7.34	7.65
	19	6.98	2.09	1.00	10.00		6.81	7.15
	20	5.20	2.72	1.00	10.00		4.98	5.42

Note. *A second way to calculate scale reliability takes into consideration the two underlying factors: This is an average of the two factors' reliabilities, which is .876.

Table 4. *EFA Eigenvalues and Variance Explained*

Factor	Initial Eigenvalues			Parallel Analysis		Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Mean of random Eigenvalues	95 percentile of random Eigenvalues	Total	% of Variance	Cumulative %	Total
1	9.289	46.446	46.446	0.377118	0.441838	8.915	44.576	44.576	7.389
2	2.087	10.433	56.879	0.314026	0.36414	1.593	7.965	52.541	8.017
3	1.070	5.351	62.230	0.261066	0.300815	0.649	3.246	55.786	5.335
4	0.992	4.959	67.189	0.221831	0.254301				
5	0.829	4.145	71.334	0.185349	0.215417				
6	0.779	3.897	75.231	0.152288	0.178079				
7	0.630	3.149	78.380	0.121088	0.145391				
8	0.586	2.932	81.312	0.088959	0.115415				
9	0.560	2.799	84.111	0.060306	0.086341				
10	0.482	2.412	86.522	0.031389	0.057957				
11	0.435	2.177	88.699	0.006237	0.02873				
12	0.398	1.988	90.686	–0.02034	–0.002315				

13	0.356	1.78	92.466	–0.048108	–0.027255
14	0.305	1.524	93.990	–0.074808	–0.055682
15	0.256	1.282	95.273	–0.100725	–0.077639
16	0.250	1.252	96.524	–0.128119	–0.107577
17	0.219	1.095	97.620	–0.155824	–0.132855
18	0.190	0.952	98.572	–0.183914	–0.160704
19	0.164	0.822	99.394	–0.217941	–0.193218
20	0.121	0.606	100.00	–0.256622	–0.228595

Note. Extraction method: Principal axis factoring. [a] When factors are correlated sums of squared loadings cannot be added to obtain a total variance.

Table 5. *EFA Item Loadings on the Three Factors with Eigenvalues Greater Than 1*

Item*	Location and component on Teimouri et al.'s (2022) scale**	Factor 1	Factor 2	Factor 3
P1) I am a diligent Spanish language learner.	1, Effort 2	**0.749***	0.022	–0.196*
P2) When it comes to Spanish, I am a hard-working learner.	3, Effort 4	**0.838***	–0.014	–0.185*
P3) Now that I have decided to learn Spanish, nothing can prevent me from reaching this goal.	5, Effort 3	0.230*	**0.682***	–0.022
P4) I will not give up learning Spanish until I master it.		0.074*	**0.822***	–0.02
P5) I have given up learning Spanish at times before and started learning it again.		0.005	–0.094	0.308*
P6) I put much time and effort into improving my Spanish language weaknesses.	9, Effort 5	**0.562***	0.263*	0.014
P7) I have overcome challenges during the process of learning Spanish.		**0.452***	0.198*	0.047
P8) I have difficulty maintaining my focus while doing my Spanish assignments.		–0.239*	0.184*	**0.473***
P9) No matter how long it takes, I will continue learning Spanish until I reach my goal.		0.006	**0.916***	0.042
P10) I will not allow anything to stop me from my progress in learning Spanish.	6, Effort 1	–0.003	**0.993***	0.136*
I1) My interests in learning Spanish change from year to year.	2, Interest 3	0.130*	-0.014	**0.627***
I2) I was obsessed with learning Spanish in the past but I have lost interest recently.	8, Interest 2	0.014	0.171	**0.792***
I3) Setbacks do not discourage me from learning Spanish.		0.171*	0.333*	–0.158*
I4) My interest in learning Spanish will not change, no matter what happens.		–0.034	**0.637***	–0.229*
I5) I have lost my interest in learning Spanish.	4, Interest 1	-0.03	–0.198*	**0.661***
I6) I am not as interested in learning Spanish as I used to be.	7, Interest 4	0.007	–0.045	**0.831***
I7) Nothing can distract me from learning Spanish.		0.082	**0.517***	–0.116*
I8) I try to keep myself motivated while learning Spanish.		**0.483***	0.257*	0.015
I9) I keep myself interested during the process of learning Spanish.		**0.517***	0.214*	–0.169*
I10) I have never lost my passion in learning Spanish.		0.175*	0.262*	–0.372*

Notes. *In column one, "P" is for *perseverance of effort*, and "I" is for *interest*, and the number following *P* or *I* refers to the item location in this study's L2 grit scale. **In column two, the first number indicates the item location in Appendix 1 in Teimouri et al.'s (2022) study. After the comma, "Effort" or "Interest" plus the number refers to the item's location in Table 3 in Teimouri et al. Bolded numbers in the last three columns represent items we retained to represent the factor based on the criterion that they loaded above absolute .4.

Factor 2 (Consistency of Interest) --> 4 Items on Interest

Likewise, people have a characteristic called *consistency of interest*, and it can be adequately measured by the items numbered 6 through 9 in the model, and we used the same 10-point Likert scale as we did with *perseverance of effort*. One's amount of consistency of interest will drive that person's response to the four items: Low responses on the four items' Likert-scale (e.g., a 1 meaning *definitely not like me* in response to "I have lost my interest in learning Spanish") will indicate high consistency of interest, which is one of two elements (or factors) of L2 grit.

Factor 1 <--> Factor 2

We hypothesize within the model that Factor 1 *Perseverance of effort* and Factor 2 C*onsistency of interest* will be related factors, but we do not have directional hypotheses based on empirical data. Teimouri et al. (2020) did not provide a correlation estimate to show the relationship between the two L2 grit components, but we hypothesize that those with high perseverance of effort will have high consistency of interest, meaning that those with high scores on Factor 1 will have low scores on Factor 2: This is because *consistency of interest* is measured negatively.

Factor 1 (Perseverance of Effort) --> L2 Grit; Factor 2 (Consistency of Interest) --> L2 Grit

We hypothesize within the model that L2 grit is composed of two distinct components, as described by Teimouri et al. (2022): As *perseverance of effort* and *consistency of interest*, with each factor contributing uniquely to a person's L2 grit. These factors may or may not be strongly related within an individual, but overall, they compose two important elements of L2 grit.

L2 Grit --> Spanish Course Grade; L2 Grit --> Overall GPA

This is a regression analysis part of the model specifying that L2 grit will be related independently to the learners' Spanish course grade and their overall GPA. We expect the relationship between L2 grit and Spanish course grade to be stronger than between L2 grit and overall GPA.

SEM Results

In Table 6, we present the maximum likelihood estimates for the 9-item model of L2 grit and our learners' L2 achievement. Table 6 shows the unstandardized and standardized estimates, along with the *p* values of the unstandardized estimates. Table S1 in the Supplemental file contains the means, standard deviations, and intercorrelations among the indicator (rectangular-shaped; see Figure 2) variables in the 2020 L2 grit model.

Based on recommendations from Kline (2011), we used the following fit indexes to assess the fit of the hypothesized model on the data: chi-square (χ^2), the χ^2 minimum divided by the degree of freedom (CMIN/DF), the root mean square error of approximation (RMSEA), the incremental fit index (IFI), and the comparative fit index (CFI) (see Table 7 for details). As written by Houghton and Jinkerson (2007), we inspected various fit statistics because "the use of multiple fit indexes is generally advisable in order to provide convergent evidence of model fit" (p. 49).

In Table 7, we also provide information on what values are desired as indicators of model fit based on reporting by Kline (2011) and Houghton and Jinkerson (2007), and we interpret the values estimated against the desired values by labeling each index in Table 7 as providing evidence for good, marginal, or poor model fit.

Figure 2. *Structural Equation Model of 2020 L2 Grit and its Influence on 2020 Achievement (N = 580).*

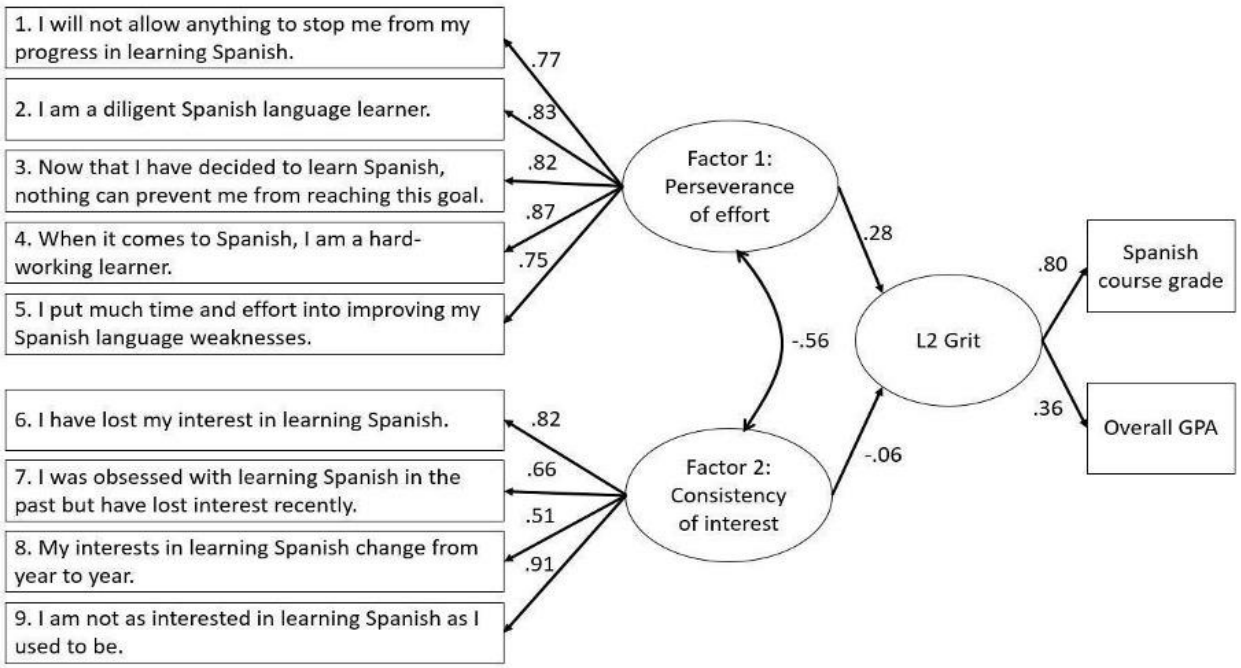

Table 6. *Maximum Likelihood Estimates (Regression Weights) for the 9-Item Model of L2 Grit and L2 Achievement*

Parameter	Unst.	SE	*p*	St.
1 (10) <-- Factor 1	1.24	0.07	<0.001	0.76
2 (1) <-- Factor 1	1.14	0.06	<0.001	0.83
3 (3) <-- Factor 1	1.35	0.07	<0.001	0.83
4 (2) <-- Factor 1	1.13	0.05	<0.001	0.87
5 (6) <-- Factor 1	1.00			0.75
6 (5) <-- Factor 2	0.8	0.04	<0.001	0.81
7 (2) <-- Factor 2	0.67	0.04	<0.001	0.66
8 (1) <-- Factor 2	0.56	0.04	<0.001	0.52
9 (6) <-- Factor 2	1.00			0.91
Factor 1 --> L2 Grit	0.08	0.02	<0.001	0.28
Factor 2 --> L2 Grit	–0.07	0.01	0.53	–0.04
L2 Grit --> Spanish Course Grade	1.00			0.81
L2 Grit --> GPA	0.68	0.26	0.01	0.36
Covariance				Corr.
Factor 1 <--> Factor 2	–2.03	0.206	<0.001	–0.55

Note. Unst. = unstandardized, SE = standard error, *p* = probability, St. = standardized, Corr. = correlation.

Table 7. *Fit Indices for the Structural Equation Model*

Model	χ^2	*df*	*p*	CMIN/*df*	RMSEA	(90% CI)	IFI Delta[2]	CFI
2020 data (*N* = 580) (Figure 2)	366.97	41	0	8.95	0.117	(.106, .128)	0.899	0.898
Good model fit	Small (close to 0)	Positive	=/<.05	=/<2.00 (1 expected)	.05–.08		>.90	>.90
Conclusion on 2020 model fit	Marginal	Good	Good	Poor	Marginal		Good	Good

Note. *df* = degrees of freedom, CMIN/*df* = chi-square minimum / *df*, RMSEA = root mean square error of approximation, IFI = incremental fit index, CFI = comparative fit index.

DISCUSSION AND CONCLUSION

When we first read Teimouri et al.'s (2022) original paper as an advance online publication in 2020, we were concerned that we could not use Teimouri et al.'s original data to trace how they derived the 9-item questionnaire to represent L2 grit out of the 20 items originally constructed. We also wondered what we would have found if we had been able to use their data to verify that they suppressed loadings that fell between –.40 and .40, that is, to check if they used absolute .40 as the criteria for item suppression. These questions drove our initial reasoning for a conceptual replication. We believe it is imperative that applied linguists publish their data (if they own them and can be appropriately anonymized) to help other researchers understand the methods used and so that they can replicate the findings.

Our first research question asked how Teimouri et al.'s (2022) model of L2 grit would fit with the data from learners of Spanish-as-a-foreign-language while including student participants beyond those majoring in the language to address one of the original paper's limitations. Generally, the results suggest that the original study's 9-item L2 grit scale measured the construct. However, we claim that it may not be strong nor accurate enough. After running an EFA on the original 20 items with the 2020 survey data, we found three factors with Eigenvalues greater than 1, instead of the original two (i.e., *perseverance of effort* and *consistency of interest*). Seven of these items corresponded to the factor that could be called *positively worded perseverance of effort* (e.g., "I am a diligent Spanish language learner"); the second element that we could call *negatively worded perseverance of effort* included another 7 items (e.g., "I will not allow anything to stop me from my progress in learning Spanish"); and the 6 remaining items encompassed the third factor that we would call *A lack of consistency of interest* (e.g., "I am not as interested in learning Spanish as I used to be"). In addition, after we ran (a second) EFA with the 9 chosen items from the L2 grit scale in Teimouri et al.'s (2022) Appendix, we found that the two factors may, in some ways, just be mirror opposites. Although some items were identified as loading on *consistency of interest* as in Teimouri et al.'s study, they actually loaded highly in our study on the factor of *perseverance of effort* as well. Because these items cross-loaded, a decision not to use them within the survey could be made. That is, cross-loaded items are sometimes seen as problematic items, and there is theoretical and statistical justification for removing them: for example, a guideline from Worthington and Whittaker (2006) is to delete a cross-loaded item if there is less than a .15 difference between its highest factor loadings.

Our findings align with previous research (Credé et al., 2017; Credé & Tynan, 2021) that focused on the original, domain-general grit scales developed by Duckworth et al. (2007) and Duckworth and Quinn (2009). Credé and Tynan (2021) explained that the 12-item and 8-item grit models presented originally (from which Teimouri et al., 2022, developed their own L2 grit scale) should not be implemented and/or adapted as they are of psychometrically poor quality. Similar to the item-wording issues we stated above, Credé and Tynan posited, "Negatively worded and positively worded items are well-known to result in artifactually distinct factors (…) and the use of these scales

therefore makes it almost impossible to correctly estimate the size of the correlation between the two constructs" (p. 39).

Furthermore, in line with the outcomes of our own SEM, Credé et al.'s (2017) meta-analysis results showed that *perseverance of effort* was a much better predictor of both performance and grade criteria than the *consistency of interest* facet. In fact, when taken to the language learning context, *consistency of interest* also seems to be a weaker predictor of L2 grit than *perseverance of effort* (e.g., Paradowski & Jelińska, 2023). Thus, Credé and colleagues (2017; 2021) expressed that the two elements of grit should be treated as two different constructs rather than a single latent trait, and that grit researchers should develop new scales with a possible future focus on p*erseverance of effort* as the most promising avenue. Our data agree with this idea (Figure 2). Another possibility we suggest is that a distinction between *perseverance of effort and a lack of consistency of interest* (since students are asked to estimate, literally, just that within themselves by responding to negatively worded items within the second construct) could be established. Moreover, it seems that applied linguistics researchers have already started to consider new ways of developing and investigating some of these constructs: For example, Alamer (2022, p. 3) refined a version of consistency of interest as an "autonomous single language interest."

In summary, our research and other research appears to suggest that the field of SLA needs to continue to examine ways to measure L2 grit and could do so through a more heightened positive psychology lens that takes a "strengths-based approach wherein one capitalizes on what one does especially well, [which] is in direct contrast to a deficiency approach that focuses on detecting and improving weaknesses" (MacIntyre et al., 2019, p. 268).

Limitations and Future Directions

With our study, we examined whether Teimouri et al.'s (2022) results from an investigation on how to measure learners' levels of grit when they are learning an L2 could be replicated with a different population (namely, L1-English learners of Spanish). In other words, we analyzed whether we could reconstruct their 9-item L2 grit scale from the original 20 L2 grit questions by following a conceptually similar data reduction plan. Post-hoc, we investigated whether their 9-item L2 grit model fit our data, mainly because we were curious to see if it would. To better examine the relationship between L2 grit and students' success in learning Spanish, we also collected language background and achievement data from the learners, including information on whether the learners were heritage language learners of Spanish. Here lies the first possible limitation of this study, since we could not look at heritage learners as a subgroup due to the small heritage learner sample size (as we indicated, we planned to do in our pre-registration; Etchart et al., 2020). We still think this is a worthy research question, as heritage language learners are individuals who have a strong cultural connection to the heritage language and may report differences in their levels of L2 grit in comparison to their non-heritage-language-learning counterparts.

Another possible limitation of this study is its "conceptual replication" nature since we in fact failed to conceptually replicate the results of the original study. Our data reduction plan from 20 to fewer items resulted in a three-factor scale with 17 items. Our study could have ended there. However, after exploring Teimouri et al.'s (2022) scale further by testing their 9-item model of L2 grit on our data using SEM, we did find that their model worked moderately well but has room for improvement. A next step would be to collect data from a new population, and perhaps provide them all 20 items again, except this time, run two or more SEMs, with at least one replicating Teimouri et al.'s model, and perhaps one replicating the 17-item model of the L2 grit scale (see Table 5) with a three-factor structure. A possible outcome could be that all models fit the data – are plausible models – but the models may differ in their fit statistics (that is, one may fit the particular data best). A very clear outcome of this study is that future scholars in the fields of applied linguistics and SLA must consider the importance of open science practices to enhance research quality, transparency, reproducibility, and accessibility that would consequently support the continuing efforts and initiatives regarding open scholarship that have already been promoted. Scale construction research needs the data to be published alongside the research, so that scale refinement and validation efforts can continue forward.

Our inability to replicate the path taken in the original study, yet a moderate fit of the 9-item L2 grit model into a different data set, may suggest that more scale-validation

studies on the L2 grit instrument are needed in the future. Whereas Teimouri et al.'s (2022) scale seems to be measuring L2 grit, our study stresses, in agreement with MacIntyre and Khajavy (2021), that questions about the scale's construction and construct representation remain. Although we see that there are a growing number of studies on L2 grit being conducted (see Table 1), we still wonder whether the original domain-general grit scale from which the language-domain-specific grit scale comes actually matches Duckworth et al.'s (2007) theory.

Finally, we believe that it would be interesting for future studies to explore the relationship between L2 grit and sustained, long-term motivation (Dörnyei & Henry, 2022) to offer further understanding on whether or not L2 grit is just another name for that well-defined construct in SLA. We speculate on this issue because we believe that if *consistency of interest* is statistically a low or non-contributing factor to the construct of L2 grit as seen in Figure 2 (and supported by previous research), the influential items remaining that comprise *perseverance of effort* may overlap tremendously with Dörnyei and Henry's (2022) theories regarding sustained, long-term motivation, which Dörnyei and Henry defined as "the maintenance or persistence of effort" (p. 90) in learning an L2. Thus, it could be that L2 grit is, conceptually, simply another name for what Dörnyei and Henry called "motivational persistence" (p. 90), which they equated with grit (p. 120), and which they defined as overlapping with grit. Sudina and Plonsky (2021a) commented on this too, in that they found L2 grit is a strong correlate with intended effort (.78) and speculated on L2 grit's overlap with student perseverance. A question is, do they overlap so much that L2 grit is just a new coinage for a construct already identified as important within SLA processes? Have L2 grit researchers been implicating long-term motivation all along?

Acknowledgments

The authors would like to thank Dr. Meagan Driver for helpful discussions and suggestions during the initial stages of this project and Dr. Wenyue (Melody) Ma for her assistance with the analysis of the data.

Authors' Contributions

Martiniano Etchart and Dr. Paula Winke participated in the design of the study. Martiniano Etchart completed the data collection. Paula Winke and Martiniano Etchart worked on data analysis. Both authors were involved in the writing of the manuscript. Paula Winke and Martiniano Etchart drafted the manuscript. Paula Winke participated in the interpretation of the results. Both authors read and approved the final manuscript.

Ethics Approval & Consent to Participate

This study was approved by the University Research Ethics Committee (IRB Study no. 5358). All participants provided written informed consent prior to enrollment and data collection in the study.

Funding

Martiniano Etchart was supported by the Research Fellowship GR100026 from the Second Language Acquisition Knowledge and Production Lab at Michigan State University. No further funding or financial support of any kind was received by the authors for this study.

REFERENCES

Alamer, A. (2022). Having a single language interest autonomously predicts L2 achievement: Addressing the predictive validity of L2 grit. *System, 108*, Article 102850. https://doi.org/10.1016/j.system.2022.102850

Alazemi, A. F. T., Jember, B., & Al-Rashidi, A. H. (2023). How to decrease Test Anxiety: A focus on academic emotion regulation, L2 grit, resilience, and self-assessment. *Language Testing in Asia*, *13*, Article 282023. https://doi.org/10.1186/s40468-023-00241-5

American Council on the Teaching of Foreign Languages. (2024). *ACTFL proficiency guidelines*. https://www.actfl.org/educator-resources/actfl-proficiency-guidelines

American Psychological Association (2020). *Publication manual of the American Psychological Association: The official guide to APA style* (7th ed.). https://doi.org/10.1037/0000165-000

Backhaus, K., Erichson, B., Gensler, S., Weiber, R., & Weiber, T. (2021). *Multivariate analysis: An application-oriented introduction.* Springer.

Banse, H., & Palacios, N. (2018). Supportive classrooms for Latino English language learners: Grit, ELL status, and the classroom context. *The Journal of Educational Research, 111*(6), 645–656. https://doi.org/10.1080/00220671.2017.1389682

Changlek, A., & Palanukulwong, T. (2015). Motivation and grit: Predictors of language learning achievement. *Veridian E-Journal, Silpakorn University (Humanities, Social Sciences and Arts), 8*(4), 23–36. https://he02.tci-thaijo.org/index.php/Veridian-E-Journal/article/view/40089

Credé, M., & Tynan, M. C. (2021). Should language acquisition researchers study "Grit"? A cautionary note and some suggestions. *Journal for the Psychology of Language Learning, 3*, 37–44. https://doi.org/10.52598/jpll/3/2/3

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

Derakhshan, A., Solhi, M., & Azari Noughabi, M. (2023) An investigation into the association between student-perceived affective teacher variables and students' L2-Grit. *Journal of Multilingual and Multicultural Development*. https://doi.org/10.1080/01434632.2023.2212644

Dörnyei, Z., & Henry, A. (2022). Accounting for long-term motivation and sustained motivated learning: Motivational currents, self-concordant vision, and persistence in language learning. In A. J. Elliot (Ed.), *Advances in motivation science* (Vol. 9, pp. 89–134). Academic Press.

Duckworth, A. (2016). *Grit: The power of passion and perseverance*. Scribner/Simon & Schuster.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology, 92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

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

Elahi-Shirvan, M., Taherian, T., & Yazdanmehr, E. (2022). L2 Grit: A longitudinal confirmatory factor analysis-curve of factors model. *Studies in Second Language Acquisition, 44*(5), 1449–1476. https://doi.org/10.1017/S0272263121000590

Etchart, M., & Winke, P. (2023). *Reexamining the construct validity of the first L2 grit test: A conceptual replication of the scale-construction processes within Teimouri, Plonsky, and Tabandeh (2022)* [Data set and codebook]. Open Science Framework. https://doi.org/10.17605/OSF.IO/9VMW5

Etchart, M., Winke, P., & Driver, M. (2020). *Measuring grit in learning Spanish* [Preregistration]. Open Science Framework. https://doi.org/10.17605/OSF.IO/U8H3E

Garnaat, S.L., & Norton, P.J. (2010). Factor structure and measurement invariance of the Yale-Brown Obsessive Compulsive Scale across four racial/ethnic groups. *Journal of Anxiety Disorders, 24*(7), 723–728. https://doi.org/10.1016/j.janxdis.2010.05.004

Hiver, P., & Al-Hoorie, A. H. (2019). *Research methods for complexity theory in applied linguistics*. Multilingual Matters.

Houghton, J. D., & Jinkerson, D. L. (2007). Constructive thought strategies and job satisfaction: A preliminary examination. *Journal of Business*

Psychology, 22, 45–53. https://doi.org/10.1007/s10869-007-9046-9

Khajavy, G., MacIntyre, P., & 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. https://doi.org/10.1017/S0272263120000480

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

Krosnick, J., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263–314). Emerald.

Language Teaching Review Panel. (2008). Replication studies in language learning and teaching: Questions and answers. *Language Teaching, 41*(1), 1–14. https://doi.org/10.1017/S0261444807004727

Lee, J. S. (2020). 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

MacIntyre, P., Gregersen, T., & Mercer, S. (2019). Setting an agenda for positive psychology in SLA: Theory, practice, and research. The *Modern Language Journal, 103*(1), 262–74. https://doi.org/10.1111/modl.12544

MacIntyre, P., & Khajavy, G. H. (2021). Grit in second language learning and teaching: Introduction to the special issue. *Journal for the Psychology of Language Learning, 3*(2), 1–6. https://doi.org/10.52598/jpll/3/2/1

Mikami, H. (2023). 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

Oxford, R., & Khajavy, G. H. (2021). Exploring grit: "Grit Linguistics" and research on domain-general grit and L2 grit. *Journal for the Psychology of Language Learning, 3*(2), 7–36. https://doi.org/10.52598/jpll/3/2/2

Paradowski, M. B., & Jelińska, M. (2023). The predictors of L2 grit and their complex interactions in online foreign language learning: Motivation, self-directed learning, autonomy, curiosity, and language mindsets. *Computer Assisted Language Learning, 37*(8), 2320–2358. https://doi.org/10.1080/09588221.2023.2192762

Patil, V. H., Singh, S. N., Mishra, S., & Donavan, D. T. (2017). *Parallel analysis engine to aid in determining number of factors to retain using R* [Computer software]. https://analytics.gonzaga.edu/parallelengine

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*(1), 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5

Sudina, E., & Plonsky, L. (2021a). Academic perseverance in foreign language learning: An investigation of language-specific grit and its conceptual correlates. *The Modern Language Journal, 105*(4), 829–857. https://doi.org/10.1111/modl.12738

Sudina, E., & Plonsky, L. (2021b). Language learning grit, achievement, and anxiety among L2 and L3 learners in Russia. *ITL – International Journal of Applied Linguistics, 172*, 161–198. https://doi.org/10.1075/itl.20001.sud

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

Teimouri, Y., Sudina, E., & Plonsky, L. (2021). On domain-specific conceptualization and measurement of grit in L2 learning. *Journal for the Psychology of Language Learning, 3*(2), 156–165. https://doi.org/10.52598/jpll/3/2/10

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

Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavior Research, 28*(3), 263–311. https://doi.org/10.1207/s15327906mbr2803_1

Worthington, R. L., & Whittaker, T. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 769–913. https://doi.org/10.1177/0011000006288127

Wu, W., Wang, Y., & Huang, R. (2023). Teachers matter: Exploring the impact of perceived teacher affective support and teacher enjoyment on L2 learners' grit and burnout. *System, 117,* Article 103096. https://doi.org/10.1016/j.system.2023.103096