Language Teaching (2026), 1–17
doi:10.1017/S0261444826101426

CAMBRIDGE UNIVERSITY PRESS

STUDY

Linguistic complexity and writing proficiency in assessing L2 Spanish writing

Pengzhan Yang[1], Wenqian Huang[1] and Guangyuan Yao[2]

[1]The University of Sydney School of Education and Social Work, Australia and [2]Central South University, China
Corresponding author: Guangyuan Yao; Email: yc37719@um.edu.mo

(Received: 11 October 2025 UTC; accepted: 9 August 2026 UTC)

Abstract
Although linguistic complexity is widely used to assess second language (L2) writing proficiency, the linguistic correlates of L2 Spanish writing proficiency remain relatively underexplored. This study addresses this gap by examining how selected lexical, syntactic, and morphological complexity indices predict L2 Spanish learners' standardized proficiency scores. We analyzed a corpus of 3,034 essays from L2 Spanish learners, utilizing their standardized placement test scores as a continuous measure of proficiency. A multiple linear regression model identified a concise set of complexity indices that effectively predicted proficiency scores, explaining a significant portion of the variance ($R^2 = .557$). Results reveal a clear proficiency-based linguistic profile: advanced writing was significantly predicted by greater phrasal complexity, sustained clausal subordination, and higher lexical sophistication. In contrast, beginners' writing was characterized by simpler syntactic strategies. These findings provide a robust, empirically derived profile of L2 Spanish writing development, offering valuable implications for pedagogical instruction, assessment rubrics, and automated evaluation tools.

Keywords: corpus linguistics; L2 Spanish; linguistic complexity; writing assessment; writing quality

1. Introduction

The assessment of second language (L2) writing is a cornerstone of applied linguistics, with linguistic complexity serving as a central and enduring construct for gauging proficiency and developmental progress for over four decades (Larsen-Freeman, 1978; Ortega, 2003; Wolfe-Quintero et al., 1998). Linguistic complexity, broadly defined as the range, elaboration, and sophistication of linguistic resources deployed in production (Bulté & Housen, 2012; Lu, 2017), is a multidimensional construct comprising syntactic, lexical, and morphological dimensions (Yoon, 2017). It is considered a key indicator of a learner's ability to manipulate the resources of a target language to achieve increasingly nuanced and effective communicative goals (Ortega, 2015).

It is necessary to distinguish linguistic complexity from two closely related constructs with which it is often conflated: writing development and perceived writing quality. Writing development refers to the underlying cognitive and linguistic trajectory of a learner's L2 acquisition as their language system matures over time. Perceived writing quality, conversely, refers to the holistic evaluation or formal score assigned to a specific text by human raters or standardized rubrics (operationalized in the present study via standardized placement test scores). While linguistic complexity—the measurable range and sophistication of the linguistic features produced—is widely used as a proxy to gauge both

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

development and quality, the three are not inherently synonymous. As previous research highlights, a developmental increase in linguistic complexity does not automatically translate to a higher perceived writing quality (Crossley & McNamara, 2014).

Historically, research in this domain focused predominantly on the syntactic dimension, relying on traditional, large-grained measures such as the mean length of T-unit (MLTU) to capture overall growth (Ortega, 2003). However, a growing body of research has challenged the adequacy of these traditional indices, critiquing them for their interpretive opacity (Norris & Ortega, 2009) and advocating for a more granular and multidimensional approach. A pivotal argument in this shift stems from corpus-based analyses demonstrating that the linguistic features of academic writing are characterized less by the clausal subordination typical of spoken discourse and more by dense, elaborated phrasal structures and sophisticated vocabulary (Biber et al., 2011). This has catalyzed the development of fine-grained indices that analyze specific syntactic constructions (Kyle & Crossley, 2018), lexical features like word frequency and diversity (Yoon, 2017), and even morphological variation (Pallotti, 2015), providing a more robust toolkit for L2 writing research.

Despite this extensive body of work, a significant limitation persists: the majority of research has focused almost exclusively on L2 English writing. This has created a rich but linguistically narrow body of knowledge, leaving our understanding of complexity in other target languages comparatively underdeveloped. A distinct lack of research exists for L2 Spanish, particularly studies that employ this modern, multidimensional framework to investigate the intricate relationship between linguistic complexity and perceived writing quality. While foundational longitudinal and cross-linguistic studies have begun to map the developmental territory of L2 Spanish (e.g., Kuiken & Vedder, 2019; Menke & Strawbridge, 2019), the field lacks a systematic investigation into which specific linguistic features—across the syntactic, lexical, and morphological domains—are most predictive of quality in L2 Spanish academic writing.

Therefore, the present study aims to address this critical gap by investigating the relationship between a comprehensive set of linguistic complexity indices and holistic quality ratings of academic essays written by L2 Spanish learners. More specifically, the study examines the multiple dimensions of lexical, syntactic, and morphological complexity to capture the features of L2 Spanish writing. At the lexical level, this study examines six indices of diversity, sophistication, and phraseological patterning, such as moving average type-token ratio (MATTR), content word frequency, verb frequency, bigram mutual information (MI) score, bigram T score, and word concreteness (Díez-Ortega & Kyle, 2024). At the syntactic level, we consider three features related to clausal subordination and phrasal elaboration, including the mean number of modifiers per noun phrase, the number of subordinate clauses, and the mean length of sentences (MLS) (Bengoetxea & Gonzalez-Dios, 2021). At the morphological level, we incorporate one measure of inflectional variation named morphological complexity index (MCI) (Brezina & Pallotti, 2019). By modeling the extent to which these dimensions contribute to variation in learners' proficiency-related writing performance, the study seeks to identify which linguistic features are salient in distinguishing higher- and lower-quality L2 Spanish writing. In this way, the study not only responds to the limited research on L2 Spanish, but also offers a more explicit account of how a multidimensional approach can be operationalized in the analysis of writing quality. The findings carry significant implications for L2 Spanish pedagogy, the design of more valid and reliable writing assessment rubrics, and the broader theoretical conceptualization of linguistic development in a typologically distinct and globally important language.

2. Literature review

2.1. Linguistic complexity as a multidimensional construct

Linguistic complexity reflects the extent to which the language produced in a task is elaborate and varied (Ellis, 2003). For decades, it has served three primary purposes in research: "to gauge proficiency, to describe performance, and to benchmark development" (Ortega, 2012, p. 128). To fully capture

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

its scope, it is essential to consider its primary dimensions: syntactic, lexical, and morphological complexity.

Syntactic complexity has historically been the most extensively studied dimension of L2 writing. Early research, drawing on first language (L1) work by Hunt (1965), operationalized complexity through global, length-based measures. The T-unit, a main clause plus any attached subordinate clauses, became the primary unit of analysis, with measures like MLTU and clauses per T-unit (C/T) becoming benchmark indices (Larsen-Freeman, 1978; Ortega, 2003). Ortega's (2003) comprehensive research synthesis solidified the value of these measures, demonstrating a generally positive and linear relationship between them and L2 proficiency, especially in college-level writing.

However, subsequent research raised critical concerns about the "interpretive opacity" of such global measures (Norris & Ortega, 2009, p. 556). An increase in MLTU, for example, could result from a variety of underlying linguistic changes—from more subordination to more phrasal elaboration—that the single score could not distinguish (Kyle & Crossley, 2018). This critique led to the formulation of a multidimensional framework for syntactic complexity that separates its distinct sub-constructs, commonly including: (1) global complexity, such as MLTU and mean length of clause (MLC), (2) subordination (e.g., C/T), (3) coordination (e.g., T-units per sentence), and (4) phrasal complexity (e.g., complex nominals per clause, CN/C) (Bulté & Housen, 2014; Lu, 2011).

A major theoretical turning point was the re-evaluation of clausal versus phrasal complexity, driven by some corpus-based research (Biber & Gray, 2010; Biber et al., 2011). These studies revealed that clausal subordination, long considered a sign of syntactic maturity, is actually more characteristic of spoken, interactive registers. In contrast, advanced academic writing across disciplines is characterized by high informational density achieved through phrasal structures, particularly complex noun phrases dense with pre- and post-modifying elements like attributive adjectives, nouns as modifiers, and prepositional phrases (Casal & Lee, 2019; Parkinson & Musgrave, 2014). This evidence suggests that as L2 writers mature academically, their development should shift from clausal to phrasal elaboration.

This insight spurred the development and validation of fine-grained indices capable of capturing specific syntactic structures rather than just their aggregate length. Tools like the L2 Syntactic Complexity Analyzer (Lu, 2010) and the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (Kyle, 2016) have enabled researchers to automatically analyze dozens of specific phrasal and clausal features. For example, Kyle and Crossley (2018) applied these fine-grained indices to TOEFL independent writing tasks and found that specific phrasal features, such as the number of modifiers per noun phrase, explained significantly more variance in human-rated writing scores than traditional clausal measures. Studies using these tools have confirmed that fine-grained phrasal indices are often more powerful predictors of writing quality in L2 English academic tasks than traditional or clausal measures (Kyle & Crossley, 2018; Zhang & Lu, 2022).

Lexical complexity, referring to the range and sophistication of vocabulary, is a crucial complement to syntactic analysis. It is typically operationalized through two main sub-constructs: lexical sophistication and lexical diversity (Read, 2000; Yoon, 2017). Lexical sophistication reflects the use of less frequent, more advanced, or contextually specific words. Common indices include average word length and, more reliably, measures of word frequency derived from large reference corpora like the Corpus of Contemporary American English (Crossley & McNamara, 2011; Yoon, 2017). The core assumption is that higher proficiency is associated with the ability to deploy a less common and more precise vocabulary. Lexical diversity (or lexical variation) measures the range of different words (types) relative to the total number of words (tokens) in a text. While the simple type-token ratio (TTR) is notoriously flawed due to its sensitivity to text length, more robust, statistically derived measures such as vocd-D, which measures the type–token relationship across random samples of varying lengths to provide a more stable estimate of lexical diversity (McCarthy & Jarvis, 2010) have become the standard in the field, providing a more valid index of a writer's vocabulary range. Empirical evidence strongly supports the reliability of these lexical measures. For instance,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Crossley et al. (2011) demonstrated that computational indices of lexical sophistication, particularly word frequency and familiarity, were highly reliable predictors of L2 writing proficiency, effectively distinguishing between different levels of learner texts. Similarly, Yoon (2017) found that lexical diversity measures were consistent and robust predictors of L2 writing quality across different essay prompts.

Morphological complexity, though less frequently studied, offers a critical "third dimension" of linguistic complexity. It refers to the variety and range of morphological forms, encompassing both inflectional (e.g., verb tense, number) and derivational (e.g., affixation) morphology (Pallotti, 2015). This dimension is particularly important as it serves as a bridge between the lexicon and the syntax. For example, the masterful use of derivational morphology is essential for creating nominalizations (e.g., investigate → investigation), a process that transforms clausal structures into the dense noun phrases characteristic of academic prose (Ortega, 2015; Ryshina-Pankova, 2015). Recent methodological advancements have produced automated tools to quantify morphological diversity using indices like the MCI (Brezina & Pallotti, 2015), allowing for its inclusion in comprehensive models of writing development (Yoon, 2017). To illustrate its predictive validity, De La Torre García et al. (2021) applied the MCI to analyze L2 Spanish writing and found that it was a significant, positive predictor of learners' Spanish proficiency levels. This highlights the reliability of morphological indices in capturing developmental stages, particularly in highly inflected languages like Spanish.

2.2. The intricate link between linguistic complexity and writing quality

The relationship between linguistic complexity and perceived writing quality is not one of simple causality. While linguistic development naturally leads to increased complexity, this development does not always align with what is valued in writing assessment for a given task. This is the central paradox of the field: "Does writing development equal writing quality?" (Crossley & McNamara, 2014, p. 66).

Crossley and McNamara's (2014) influential study revealed a disassociation between developmental gains and rated quality in L2 English essays. Over a semester, learners' writing evolved to include more phrasal and nominal structures characteristic of academic writing. However, texts featuring greater clausal complexity (e.g., dependent clauses) tended to receive higher scores, although this feature is more aligned with spoken discourse. This suggests that expectations underlying writing assessment, potentially shaped by task, genre, or implicit biases, can reward linguistic profiles that diverge from established norms of academic writing.

This paradox underscores the critical role of mediating variables in the complexity–quality relationship. The predictive power of complexity indices is highly genre-dependent. Zhang and Lu (2022) found that distinct sets of fine-grained syntactic indices predicted writing quality in application letters versus argumentative essays, reflecting their different communicative functions. Similarly, Qin and Uccelli (2016) found that MLC was a significant quality predictor for argumentative essays but not for narratives. The topic of a writing task can influence the cognitive demands placed on a writer, thereby affecting their linguistic output. Yang et al. (2015) demonstrated that different argumentative essay topics elicited different syntactic complexity profiles, which in turn correlated differently with quality scores. The relevance of specific complexity indices changes as learners develop their L2. Features that distinguish beginning from intermediate writers (e.g., clausal coordination) may cease to be relevant at advanced levels, where phrasal elaboration and lexical sophistication become more powerful discriminators (Lu, 2011; Yoon, 2017).

While complexity is a primary focus, it cannot be fully understood in isolation from accuracy and fluency (CAF) (Skehan, 2009). The CAF framework posits that learners have limited attentional resources, leading to potential trade-offs where a focus on one dimension may come at the expense of another (Pallotti, 2009). A learner attempting highly complex syntax, for instance, may produce more errors, thus decreasing accuracy.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Recent research on L2 Spanish writing suggests that at certain proficiency thresholds, accuracy may become a more salient differentiator than complexity. In a study comparing intermediate-high and advanced-low L2 Spanish writers, Mangold (2025) found no significant differences in measures of lexical or syntactic complexity. Instead, the most powerful distinguishing feature was accuracy, specifically the number of errors per 100 words (E/100). The advanced-low group produced significantly fewer errors than the intermediate-high group. This finding suggests that learners at this crucial transition point may be consolidating their grammatical knowledge and prioritizing accuracy over further complexity (Byrnes et al., 2010). This aligns with the concept of developmental plateaus, where learners consolidate existing knowledge before pushing into new linguistic territory (Restrepo-Ramos, 2021). The reliability of accuracy measures like error-free T-units and E/100 has been well-established in L2 research (Larsen-Freeman, 1978; Polio & Shea, 2014), and Mangold's (2025) study confirms their particular relevance for distinguishing degrees of advancement in L2 Spanish.

The body of research on linguistic complexity in L2 Spanish, while growing, remains significantly smaller than that for L2 English. Existing studies provide a crucial starting point. The longitudinal case study by Menke and Strawbridge (2019) of L2 Spanish majors found significant growth in length-based measures, such as MLTU, MLC, and mean length of noun phrase, confirming a developmental trend toward greater syntactic elaboration, especially within the noun phrase. Critically, however, they observed little change in inter-clausal measures like C/T, suggesting that advanced learners of Spanish, much like their English-learning counterparts, may reach a plateau in clausal subordination and shift developmental resources toward phrasal complexity.

Cross-linguistic research further highlights how the typological features of Spanish must be considered. Kuiken and Vedder (2019) highlight that structural and stylistic preferences differ across languages, impacting how complexity is realized. For instance, Romance languages like Spanish often favor post-nominal modification (un libro rojo), contrasting with the pre-nominal modification common in Germanic languages like English (a red book). Furthermore, Spanish frequently uses prepositional phrases where English might use adjectival or nominal pre-modifiers (e.g., casa de madera for "wooden house"), a structural feature that likely influences the development and measurement of phrasal complexity (Menke & Strawbridge, 2019). These structural and stylistic differences underscore the need for a language-specific investigation rather than a simple application of models derived from L2 English.

2.3. The present study and research questions

A critical gap remains in our understanding of which specific combination of syntactic, lexical, and morphological features is most predictive of L2 Spanish proficiency. To address this, the present study employs a multidimensional framework. Rather than testing all conceivable measures cited in previous literature, we selected a focused battery of metrics across three dimensions (lexical, syntactic, and morphological) based on two primary criteria. First, we selected indices that have recently demonstrated strong predictive potential in L2 research but remain under-explored in Spanish (e.g., fine-grained phrasal modifiers, MATTR, and the MCI). Second, the selection was constrained by the availability of highly reliable, state-of-the-art automated Natural Language Processing (NLP) tools developed specifically for Spanish (e.g., TAALES_ES, MultiAzterTest).

By utilizing these carefully selected metrics, the present study is designed to answer the following research questions and test their corresponding hypotheses:

Hypothesis 1 (H1): Based on the robust findings in L2 English research (e.g., Kyle & Crossley, 2018; Lu, 2017), we hypothesize that a multidimensional regression model incorporating selected syntactic, lexical, and morphological metrics will be able to accurately predict a significant portion of the variance in continuous L2 Spanish proficiency scores.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Hypothesis 2 (H2): Drawing on previous L2 Spanish and general developmental research, we predict that fine-grained phrasal complexity (e.g., noun phrase modifiers) (Menke & Strawbridge, 2019), lexical sophistication (e.g., MATTR and lower content word frequency) (Díez-Ortega & Kyle, 2024), and morphological diversity (e.g., MCI) (De La Torre García et al., 2021) will emerge as significant positive predictors, reflecting the multidimensional nature of advanced L2 academic writing.

Based on these hypotheses, this study addresses the extent to which the selected linguistic complexity indices predict L2 Spanish learners' standardized proficiency scores. This study is guided by the following two research questions:

RQ1: Which broader dimensions of linguistic complexity—lexical, syntactic, or morphological complexity—serve as stronger predictors of L2 Spanish writing proficiency?

RQ2: Which selected linguistic complexity measures significantly predict L2 Spanish writing proficiency, and in what direction?

3. Method

3.1. Corpus

In this study, we used texts from the Corpus Escrito del Español como L2 (L2 Spanish Written Corpus; CEDEL2) version 2 (Lozano, 2022). Following established learner corpus design principles (Granger, 2009), the corpus was compiled based on controlled elicitation tasks with the information on students' standardized assessment scores and background variables (Lozano, 2022). These features make the corpus suitable for examining the relationship between linguistic complexity measures and proficiency while accounting for internal variation. Specifically, CEDEL2 included essays on 14 topics, which can be further classified as descriptive (n = 2,847) or argumentative prompts (n = 187) (see Table 1). Additionally, this corpus provides students' scores on The University of Wisconsin College-Level Placement Test of Spanish Grammar (0–43), which were used in this study as an indicator of proficiency. Information about learners' L1 backgrounds is also available. Given the controlled task design and rich learner background information, this corpus enables a more precise examination of the relationship between linguistic complexity and L2 Spanish proficiency. The inclusion of relevant covariates (e.g., L1 and task type) can help isolate the predictive roles of linguistic features from other sources of variation, strengthening the robustness of the findings. In the present study, all 3,034 texts across the 14 writing topics were included in order to maximize sample size and to examine whether a range of linguistic complexity indices could predict proficiency. The students' proficiency scores (M = 30.302, SD = 9.359) ranged from 2 to 43, indicating that the sample represented learners from basic to advanced proficiency levels. As previous research has shown that task type can affect linguistic complexity metrics (e.g., Yang et al., 2015; Yoon, 2017) and background variables may contribute to variation, task type and L1 were used as covariates in this study to reduce the potential confounding effects of task variation and learner background.

3.2. Linguistic complexity indices

We analyzed the linguistic complexity indices using three tools: TAALES_ES[1] (Díez-Ortega & Kyle, 2024), MultiAzterTest[2] (Bengoetxea & Gonzalez-Dios, 2021), and Morpho complexity tool[3] (Brezina & Pallotti, 2019). TAALES_ES generated a range of lexical complexity results with reference to a subset of the Spanish version of the Corpus of the Web (ESCOW) corpus comprising 450 million words (Schäfer, 2015). The MultiAzterTest and Morpho complexity tool were used for computing syntactic and morphological complexity measures, respectively. Table 2 presents the lexical, syntactic, and morphological indices used in this study.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Table 1. Writing topics in CEDEL2

Topic	Task type	Numberof texts
1. Describe the region where you live.	Descriptive	457
2. Talk about a famous person.	Descriptive	262
3. Describe a film you have recently seen.	Descriptive	133
4. What did you do last year during your holidays?	Descriptive	306
5. Which are your plans for the future?	Descriptive	237
6. Describe a trip you have recently made.	Descriptive	206
7. Narrate an experience you have lived.	Descriptive	56
8. Talk about the problem of terrorism in the world.	Argumentative	19
9. What do you think about the anti-tobacco law?	Argumentative	36
10. Do you think gay couples have the right to get married and adopt children?	Argumentative	81
11. Do you think marihuana should be legalized?	Argumentative	19
12. Analyze the main aspects of immigration.	Argumentative	32
13. Describe the frog story.	Descriptive	150
14. Describe the Charles Chaplin silent video clip.	Descriptive	1,040

Table 2. Linguistic complexity measures

Construct	Indices	Definition
Lexical complexity	MATTR	The MATTR based on a 50-word window.
	Content word frequency	The average frequency of content words.
	Verb frequency	The average frequency for verbs.
	Bigram MI score	The average MI score for bigrams.
	Bigram T score	The average T score for bigrams.
	Word concreteness	The average concreteness score for content words.
Syntactic complexity	The mean number of modifiers per noun phrase	The mean number of modifiers per noun phrase.
	The number of subordinate clauses	The number of subordinate clauses in the text.
	The MLS	The mean number of words in each sentence in the text.
Morphological complexity	MCI	The diversity of verbal inflectional types in a text.

3.2.1. Lexical complexity

This study examined six lexical complexity measures obtained from TAALES_ES. MATTR was calculated by averaging the TTR scores for all 50-word windows of each writing sample. More proficient L2 Spanish users are expected to get higher MATTR scores (Zenker & Kyle, 2021). Content word frequency was measured based on the mean frequency score of all adjectives, adverbs, nouns, and verbs with reference to the average content word frequency scores on the ESCOW corpus (Schäfer, 2015). More proficient L2 Spanish learners tend to display more sophisticated, lower-frequency content words (Díez-Ortega & Kyle, 2024). The mean frequency for verbs was also computed based on the ESCOW corpus to measure learners' use of verb inflection and subjunctive mood, which may

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

reflect the development of L2 Spanish users (Schnur & Rubio, 2021). For instance, less frequent verbs that indicate past tense conjugations (e.g., desestimé; I dismissed), present subjunctive forms (e.g., comprometa; were I to compromise), or future indicative forms (e.g., describiré; I will describe; describirán; they will describe) are considered more sophisticated. The higher MI score for bigrams shows learners' use of more exclusive word combinations (e.g., inversamente proporcional; inversely proportional) than those less exclusive word combinations (e.g., solo sé; I only know), while the Bigram T scores represent the frequently occurring word combinations (e.g., muy bueno; very good). Previous research in English has indicated that more advanced L2 learners tend to use bigrams that earn higher MI or T scores (e.g., Bestgen & Granger, 2014; Eguchi & Kyle, 2020; Garner et al., 2019; Paquot, 2019). Word concreteness scores refer to the degree to which a word refers to a perceptible entity. For example, abeja (bee) and manzana (apple) are related to higher concreteness scores, while salud (health) and suerte (luck) earn lower scores. In TAALES_ES, this measure is calculated based on Guasch et al.'s (2016) concreteness scores. Crossley and Skalicky (2019) highlight that more proficient learners generally get lower concreteness scores.

3.2.2. Syntactic complexity

Concerning syntactic complexity, we gained three metrics, including the mean number of modifiers per noun phrase, the number of subordinate clauses, and the mean sentence length, from MultiAzterTest. These measures have been examined in previous studies (e.g., Casal & Lee, 2019; Crossley & McNamara, 2014; Imaz Agirre et al., 2025). In L2 Spanish writing, for instance, Imaz Agirre et al. (2025) explored learners' changes in these three indices. They reported that both the mean number of modifiers per noun phrase and mean sentence length were correlated with the development of students who received instruction targeting academic language use.

3.2.3. Morphological complexity

As for morphological complexity, Brezina and Pallotti's (2019) MCI was used in this study. MCI captures variation in morphological forms by measuring the diversity of inflectional morphemes attached to a base form. As a type-based measure, it controls for text length effects by counting the number of distinct morphological exponents across the subsample segments of inflected verb forms. The calculation of this measure incorporates within-segment variety and between-segment variety. With a segment size of 10 verb tokens, MCI scores range from 0 to 19, which indicates the highest inflectional diversity, based on the formula: (within-subject variety + between-subject variety/2) − 1 (Brezina & Pallotti, 2019). The relationship between this metric and language proficiency has been examined (e.g., De la Torre García et al., 2021). De la Torre García et al. (2021) found that MCI was a significant positive predictor of L2 learners' Spanish proficiency level.

3.3. Linguistic analysis

In order to address the research questions, we conducted two statistical analyses of multiple regression, with L1 and task type used as covariates. A multiple regression was first conducted to determine the amount of variance in students' scores that could be explained by the linguistic complexity indices and to identify statistically significant indices. The MuMIn::dredge() package (Consolini & Kyle, 2024) was then utilized to explore whether a model with fewer indices would yield comparable predictive performance and to identify indices that were retained in the best-fitting models. To evaluate the predictive stability and generalizability of both models, the ten-fold cross-validation (10-fold CV) was conducted for both multiple regression models. In the procedure of 10-fold CV, 90% of the data were used to create a model, which was subsequently tested on the remaining 10% of the data. The procedure was repeated until all data had been used as the test set, with the averaged results from the ten folds reported (Tabachnick & Fidell, 2014).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

We first checked the normal distribution of the data using histograms for each variable. Afterward, scatter plots were employed to visualize the linearity of each variable in relation to the placement scores. After data screening, MLS was removed due to its violation of normal distribution and the presence of influential outliers, while other variables were retained for subsequent data analysis. Following the reviewer's suggestion, we further examined a log-transformed version of MLS; however, the transformed measure showed substantial collinearity with the number of subordinate clauses, which provides a more fine-grained indicator of syntactic complexity. Therefore, to avoid redundancy and preserve model stability, we retained the number of subordinate clauses and excluded MLS from the final model. The correlation between the remaining indices and scores was then examined (see Table 3). Most indices demonstrated significant relationships with scores ($|r| > .10$, $p < .05$). We also checked the multicollinearity of retained variables, with no retained indices showing strong intercorrelations ($|r| > .70$) (Tabachnick & Fidell, 2014). After the variables of MLS and verb frequency were discarded, a multiple regression was first conducted to investigate the relationship between the scores and eight indices. A 10-fold CV procedure was further employed to examine the generalizability of the model in predicting scores. Consequently, the MuMIn::dredge() function was used to identify the optimal model, and the calc.relimp() function in the relaimpo package (Grömping, 2006) was used to calculate the relative importance of the indices in the model. We also employed 10-fold CV to assess the generalizability of the final model's predictive performance.

4. Results

A multiple regression model was first conducted with eight indices to examine the extent to which linguistic complexity indices explained the variance in L2 Spanish learners' placement test scores. The overall model was significant, with $F(19, 3014) = 199.4$, $p < .001$, explaining 55.7% of the variance in proficiency scores ($R^2 = .557$, adjusted $R^2 = .554$). Table 4 provides an overview of the multiple regression model, in which MATTR, content word frequency, bigram MI score, word concreteness, the mean number of modifiers per noun phrase, the number of subordinate clauses, and MCI were significant predictors of proficiency scores ($p < .05$). However, bigram T score did not show an independent effect ($p = .516$). A follow-up 10-fold CV of the regression model yielded a mean R^2 of .550 Root Mean Squared Error (RMSE) = 6.261, Mean Absolute Error (MAE) = 4.766, indicating that the model is stable and generalizable across the data.

Based on the full model, we further identified an optimal multiple regression model using the MuMIn::dredge() function while retaining L1 and task type as covariates. The final model retained seven predictors, excluding bigram T score. The optimal model was statistically significant, explaining 55.7% of the variance in proficiency scores ($R^2 = .557$, adjusted $R^2 = .554$), with $F(18, 3015) = 210.5$, $p < .001$. Table 5 shows the results of the multiple regression model comprising the retained seven indices. 33.1% of the explained variance can be attributed to MCI. The mean number of modifiers per noun phrase, the number of subordinate clauses, and MATTR accounted for 18.8%, 14.8%, and 13.6% of the explained variance in scores, respectively. By contrast, three lexical indices, including Bigram MI score, content word frequency, and word concreteness, contributed relatively small proportions of the explained variance, ranging from 1.6% to 3.0%.

In terms of the direction of these effects, MCI, the mean number of modifiers per noun phrase, the number of subordinate clauses, MATTR, Bigram MI score, and word concreteness were positive predictors of proficiency scores, indicating that higher-scoring learners tended to produce texts with greater morphological diversity, more elaborated noun phrases, more frequent clausal subordination, higher lexical diversity, stronger word associations, and more concrete lexical choices. By contrast, content word frequency was a negative predictor, suggesting that higher proficiency scores were associated with the use of less frequent and therefore more lexically sophisticated content words.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Table 3. The correlation analysis for variables

	1	2	3	4	5	6	7	8	9	10
1. Proficiency score	1									
2. MATTR	.432***	1								
3. Content word frequency	.186***	.039*	1							
4. Verb frequency	−.076***	−.114***	.255***	1						
5. Bigram MI score	.232***	.352***	.049**	.056**	1					
6. Bigram T score	.231***	.090***	−.013	−.018	.442***	1				
7. Word concreteness	.137***	−.060***	−.043*	−.060***	.090***	.095***	1			
8. The mean number of modifiers per noun phrase	.472***	.226***	.310***	−.085***	−.010	.238***	−.002	1		
9. The number of subordinate clauses	.483***	.327***	.400***	.045*	.236***	.127***	.052**	.298***	1	
10. MCI	.610***	.375***	.244***	.027	.216***	.182***	.122***	.458***	.429***	1

Note: *$p < .05$, **$p < .01$, ***$p < .001$.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Table 4. Results of the full multiple regression model

	Estimate	SE	*t*	*p*
(Intercept)	4.972	2.399	2.072	.038
MATTR	20.300	1.736	11.696	<.001
Content word frequency	−0.699	0.185	−3.773	<.001
Bigram MI score	0.980	0.404	2.424	.015
Bigram T score	0.001	0.002	0.649	.516
Word concreteness	0.230	0.102	2.260	.024
The mean number of modifiers per noun phrase	8.734	0.543	16.085	<.001
The number of subordinate clauses	0.041	0.005	8.427	<.001
MCI	0.868	0.037	23.563	<.001

Note: L1 and task type were included as covariates in the model.

Table 5. Results of the optimal multiple regression model

	Relative importance	Estimate	SE	*t*	*p*
(Intercept)		5.182	2.377	2.180	.029
Bigram MI score	.030	1.107	0.354	3.124	.002
Content word frequency	.016	−0.715	0.184	−3.896	<.001
MCI	.331	0.869	0.037	23.610	<.001
The mean number of modifiers per noun phrase	.188	8.847	0.515	17.194	<.001
The number of subordinate clauses	.148	0.040	0.005	8.405	<.001
Word concreteness	.016	0.229	0.102	2.253	.024
MATTR	.136	20.127	1.715	11.736	<.001

Note: L1 and task type were included as covariates in the model.

A 10-fold CV was conducted to evaluate the stability and generalizability of the optimal regression model. The model showed consistent predictive performance across folds, with a mean R^2 of .550, indicating that it explained approximately 55.0% of the variance in proficiency scores on average (RMSE = 6.261, MAE = 4.767). Overall, the performance of the optimal model in 10-fold CV closely matched its performance on the full data, suggesting good generalizability across the dataset.

5. Discussion

This study investigated the relationship between linguistic complexity and learners' L2 Spanish proficiency, aiming to identify the linguistic correlates of proficiency and to validate a multidimensional predictive model. The results confirm that holistic proficiency ratings have a robust and measurable empirical basis in the linguistic features of a text (Casal & Lee, 2019; Lu, 2017; Yoon, 2017). The key findings reveal that a multidimensional linear regression model incorporating syntactic, lexical, and morphological indices can predict continuous L2 Spanish proficiency scores with a high degree of reliability. Furthermore, the specific linguistic features retained in the optimal model reveal a clear predictive profile characterized by greater morphological diversity, dual mastery of clausal and phrasal syntactic complexity, and increasing lexical sophistication (Ortega, 2015).

The first research question asked how accurately a model based on linguistic complexity could predict L2 Spanish writing proficiency. The strong performance of the optimal multiple regression model—explaining over 55% of the variance in learners' placement test scores ($R^2 = .557$)—provides a clear answer. This high level of predictive stability, supported by robust 10-fold CV results (mean

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

$R^2 = .550$, MAE = 4.767), validates the fundamental premise that the construct of proficiency in writing is not an abstract intuition but is deeply grounded in observable, quantifiable linguistic patterns (Kyle & Crossley, 2018; Lu, 2017; Zhang & Lu, 2022). The model's successful generalization from the training set to the test set further underscores its stability and potential for broader application in writing assessment. It is crucial to acknowledge that linguistic complexity is inherently sensitive to task characteristics and genre (e.g., Yang et al., 2015; Yoon, 2017). Recognizing this, the present study explicitly controlled for task type (descriptive vs. argumentative) as a covariate in the multiple regression models. By doing so, the optimal model successfully isolated the variance attributable to L2 proficiency from the confounding effects of task-specific linguistic demands. Therefore, the seven core indices identified—such as MCI and noun phrase modification—represent robust, cross-task indicators of L2 Spanish writing development, rather than artifacts of a specific writing prompt. Furthermore, it is well-documented that learner background variables can significantly impact L2 writing complexity. To ensure the robustness of our findings, the learners' L1 was explicitly included as a covariate in our multiple regression models, alongside task type. This analytical step helped isolate the relationship between linguistic complexity and L2 Spanish proficiency from the potential cross-linguistic influence of the writers' native languages. Therefore, the developmental trajectory identified in this study reflects changes driven primarily by proficiency rather than L1 transfer.

The second research question sought to identify which specific linguistic features were most predictive of proficiency. The optimal model's retention of morphological, syntactic, and lexical indices aligns with the widely accepted view that linguistic complexity is not a monolithic construct but a system of interacting dimensions (Bulté & Housen, 2014; Norris & Ortega, 2009).

In the morphological domain, the MCI emerged as the single most powerful predictor, accounting for roughly a third (33.1%) of the explained variance in proficiency scores. This underscores the critical importance of inflectional diversity in L2 Spanish. Unlike English, Spanish is a highly inflected language where verbal morphology encodes tense, aspect, and mood. The strong predictive power of MCI confirms that the ability to deploy a wide variety of verbal inflections is a primary hallmark of advanced L2 Spanish writing, corroborating recent findings by De La Torre García et al. (2021) and highlighting morphology as an essential "third dimension" of complexity (Pallotti, 2015).

In the syntactic domain, the results reveal that as L2 Spanish learners become more proficient, they demonstrate greater use of both phrasal elaboration and clausal subordination. The significant predictive power of the mean number of modifiers per noun phrase confirms the importance of phrasal complexity, aligning with the large body of research identifying the noun phrase as a primary site of development in academic and advanced writing (Biber et al., 2011; Casal & Lee, 2019; Parkinson & Musgrave, 2014), and corroborating similar longitudinal findings in L2 Spanish (Menke & Strawbridge, 2019). Simultaneously, the positive predictive power of the number of subordinate clauses indicates that clausal subordination remains a key differentiator of proficiency in this corpus. This suggests that the developmental trajectory for Spanish may involve a more sustained, parallel growth in both phrasal and clausal complexity, reflecting structural and stylistic typological preferences in Romance languages (Kuiken & Vedder, 2019).

Crucially, the prominence of these specific predictors provides novel insights into L2 complexity across typologically distinct languages. Unlike the vast majority of L2 English models—where syntactic measures typically dominate—our findings reveal that in L2 Spanish, morphological diversity (MCI) is the paramount predictor. This directly reflects the typological nature of Spanish as a highly inflected, pro-drop (null-subject) language. Because explicit subject pronouns are optional and often omitted in Spanish, the communicative and grammatical burden is heavily shifted onto verbal inflection. Therefore, mastering the rich inflectional system is a distinctly Spanish-centric

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

developmental milestone. Furthermore, while L2 English research often observes a strict developmental shift away from clausal subordination toward phrasal elaboration (Biber et al., 2011), our L2 Spanish model indicates a concurrent, sustained reliance on both. This underscores the stylistic and structural preferences of Romance languages, which frequently utilize post-nominal prepositional phrases and relative clauses for elaboration, contrasting with the pre-nominal modification typical of Germanic languages (Kuiken & Vedder, 2019). Ultimately, these findings demonstrate that models of L2 writing development must be theoretically anchored in the specific typological features of the target language.

In the lexical domain, the retention of four distinct indices highlights the multidimensional nature of vocabulary development. The strong positive contribution of MATTR confirms a universal marker of advanced proficiency: a larger and more varied vocabulary (Crossley et al., 2011; Read, 2000; Yoon, 2017). Furthermore, the positive correlation with the Bigram MI score indicates that higher proficiency is associated with the use of more exclusive and strongly associated word combinations (phraseological competence), supporting findings in L2 English (Bestgen & Granger, 2014; Paquot, 2019). Crucially, content word frequency emerged as a negative predictor, confirming that as proficiency increases, learners shift away from highly frequent, generic words toward more sophisticated, lower-frequency vocabulary (Díez-Ortega & Kyle, 2024). Although word concreteness was a positive predictor in this model—contrary to some expectations that advanced writing becomes more abstract (Crossley & Skalicky, 2019)—this may be partially attributed to the predominantly descriptive nature of the corpus tasks, which heavily elicit concrete nouns and adjectives. Accordingly, the predictive value of word concreteness should be interpreted with caution, as it is likely to be highly sensitive to prompt genre and should not be generalized across writing tasks or incorporated into automated writing evaluation systems without genre-specific validation.

This study's findings also provide a valuable point of synthesis with previous research on L2 Spanish. Mangold (2025) found that at a very specific threshold (intermediate-high to advanced-low), accuracy was a more powerful predictor than complexity. The present study, examining a broader proficiency continuum (0–43 scores), demonstrates that complexity measures are indeed highly predictive across wider developmental stages. This aligns with models positing trade-offs between CAF, suggesting that while the overarching journey from beginner to advanced is marked by continuous complexity, localized stages may involve temporary plateaus where learners prioritize accuracy to consolidate newly acquired complex structures (Pallotti, 2009; Skehan, 2009).

While the current findings are primarily predictive, they lay valuable groundwork for future research into the teaching and assessment of L2 Spanish writing. Pedagogically, the linguistic profile identified here suggests that future instructional interventions could explore the efficacy of explicitly targeting these features—such as building dense noun phrases and mastering morphological diversity. However, longitudinal, classroom-based research is needed to validate whether teaching these specific structures directly translates to improved writing proficiency across different learner populations. For assessment, this study highlights the potential value of moving beyond vague rubrics toward multidimensional criteria. Nevertheless, before these indices can be widely applied, future research must validate them against holistic human ratings across diverse genres and task conditions to fully determine their viability for localized rubrics or the development of automated writing evaluation tools.

6. Conclusion

This study set out to address a significant gap in the literature by investigating the linguistic features that predict learners' L2 Spanish writing proficiency. By applying a multidimensional computational analysis to a large corpus of learner essays, we have demonstrated that L2 Spanish proficiency can be predicted with a high degree of accuracy based on a combination of syntactic, lexical, and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

morphological indices. The findings reveal a clear linguistic profile associated with L2 Spanish proficiency, where higher-scoring texts are characterized by a sophisticated command of phrasal elaboration, clausal subordination, and lexical diversity, in contrast to the simpler syntactic strategies found at lower proficiency levels.

The primary contribution of this research is the provision of large-scale, empirical evidence that specifies the linguistic correlates of proficiency in L2 Spanish writing, moving the field beyond its traditional focus on L2 English and building upon foundational work in L2 Spanish (e.g., Kuiken & Vedder, 2019; Mangold, 2025; Menke & Strawbridge, 2019). The results both confirm universal developmental patterns (e.g., the importance of lexical diversity) and highlight language-specific trends (e.g., the sustained importance of clausal subordination alongside phrasal growth).

This study is not without its limitations. First, its cross-sectional design allows us to infer development but not to track it directly within individual writers. Second, the proficiency levels were determined by a placement test rather than a holistic score assigned to each essay. Third, although we explicitly controlled for task type and L1 as covariates, we acknowledge that genre and task conditions may still interact with linguistic complexity in subtle ways not fully captured by this broad control. Similarly, other crucial background variables—such as learners' specific instructional contexts and their overall exposure to Spanish—were not incorporated into our models due to data availability constraints in the CEDEL2 corpus. These unmeasured factors may influence the generalizability of our findings. Investigations into different genres of L2 Spanish writing (e.g., narrative, professional) would illuminate how complexity features vary by communicative context (cf. Yang et al., 2015; Zhang & Lu, 2022). Finally, while computational analysis offers immense power and consistency, it is not immune to parsing errors. Future research should build on these findings through longitudinal studies to confirm the developmental pathways identified here (e.g., Bulté & Housen, 2018). Most importantly, future work should integrate measures of accuracy alongside complexity to create a more comprehensive CAF model of L2 Spanish writing development, exploring the trade-offs and interactions between these critical dimensions at different proficiency stages (Norris & Ortega, 2009; Skehan, 2009). Ultimately, this study affirms that as L2 Spanish learners develop, their writing becomes measurably more complex in systematic ways. Understanding these patterns is essential for creating more effective pedagogies and more valid and reliable assessments that can support learners on their path to advanced proficiency.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0261444826101426.

Data availability statement. The relevant materials supporting the analyses reported in this study (R scripts) are provided as Supplementary Materials.

Notes

1. https://github.com/LCR-ADS-Lab/TAALES_ES.
2. https://github.com/kepaxabier/MultiAzterTest.
3. http://corpora.lancs.ac.uk/vocab/analyse_morph.php.

References

Bengoetxea, K., & Gonzalez-Dios, I. (2021). Multiaztertest: A multilingual analyzer on multiple levels of language for readability assessment. *arXiv preprint*. https://doi.org/10.48550/arXiv.2109.04870

Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing*, *26*, 28–41. https://doi.org/10.1016/j.jslw.2014.09.004

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*, *9*(1), 2–20. https://doi.org/10.1016/j.jeap.2010.01.001

Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45*(1), 5–35. https://doi.org/10.5054/tq.2011.244483

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Brezina, V., & Pallotti, G. (2015). Morphological complexity tool. Available from: http://corpora.lancs.ac.uk/vocab/analyse_morph.php

Brezina, V., & Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second Language Research*, *35*(1), 99–120. https://doi.org/10.1177/0267658316643125

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Investigating complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins Pub. Co. https://doi.org/10.1075/lllt.32.02bul

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, *26*, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

Bulté, B., & Housen, A. (2018). Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics*, *28*(1), 147–164. https://doi.org/10.1111/ijal.12196

Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *Modern Language Journal*, *94*(s1), i–235. http://www.jstor.org/stable/40985261

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing*, *44*, 51–62. https://doi.org/10.1016/j.jslw.2019.03.005

Consolini, C. H., & Kyle, K. (2024). The relationship between L2 Spanish proficiency and features of written lexical and lexicogrammatical use. *Applied Linguistics*, *45*(6), 992–1009. https://doi.org/10.1093/applin/amae032

Crossley, S. A., & McNamara, D. S. (2011). Understanding expert ratings of essay quality: Coh-Metrix analyses of first and second language writing. *International Journal of Continuing Engineering Education and Life Long Learning*, *21*(2-3), 170–191. https://doi.org/10.1504/IJCEELL.2011.040197

Crossley, S. A., & McNamara, D. S. (2014). Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, *26*, 66–79. https://doi.org/10.1016/j.jslw.2014.09.006

Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, *28*(4), 561–580. https://doi.org/10.1177/0265532210378031

Crossley, S. A., & Skalicky, S. (2019). Examining lexical development in second language learners: An approximate replication of Salsbury, Crossley & McNamara (2011). *Language Teaching*, *52*(3), 385–405. https://doi.org/10.1017/S0261444817000362

De La Torre García, N., Ainciburu, M. C., & Buyse, K. (2021). Morphological complexity and rated writing proficiency: The case of verbal inflectional diversity in L2 Spanish. *ITL-International Journal of Applied Linguistics*, *172*(2), 290–318. https://doi.org/10.1075/itl.20009.del

Díez-Ortega, M., & Kyle, K. (2024). Measuring the development of lexical richness of L2 Spanish: A longitudinal learner corpus study. *Studies in Second Language Acquisition*, *46*(1), 169–199. https://doi.org/10.1017/S0272263123000384

Eguchi, M., & Kyle, K. (2020). Continuing to explore the multidimensional nature of lexical sophistication: The case of oral proficiency interviews. *Modern Language Journal*, *104*(2), 381–400. https://doi.org/10.1111/modl.12637

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Garner, J., Crossley, S. A., & Kyle, K. (2019). N-gram measures and L2 writing proficiency. *System*, *80*, 176–187. https://doi.org/10.1016/j.system.2018.12.001

Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13–32). John Benjamins Publishing Company. https://doi.org/10.1075/scl.33

Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software*, *17*(1), 1–27. https://doi.org/10.18637/jss.v017.i01

Guasch, M., Ferré, P., & Fraga, I. (2016). Spanish norms for affective and lexico-semantic variables for 1,400 words. *Behavior Research Methods*, *48*(4), 1358–1369. https://doi.org/10.3758/s13428-015-0684-y

Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. The National Council of Teachers of English.

Imaz Agirre, A., Arias-Hermoso, R., & Ipiña, N. (2025). The effect of an intervention focused on academic language on CAF measures in the multilingual writing of secondary students. *International Review of Applied Linguistics in Language Teaching*, *63*(2), 1373–1396. https://doi.org/10.1515/iral-2023-0137

Kuiken, F., & Vedder, I. (2019). Syntactic complexity across proficiency and languages: L2 and L1 writing in Dutch, Italian and Spanish. *International Journal of Applied Linguistics*, *29*(2), 192–210. https://doi.org/10.1111/ijal.12256

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication* [Unpublished doctoral dissertation]. Georgia State University.

Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in l2 writing using fine-grained clausal and phrasal indices. *Modern Language Journal*, *102*(2), 333–349. https://doi.org/10.1111/modl.12468

Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, *12*(4), 439–448. https://doi.org/10.2307/3586142

Lozano, C. (2022). CEDEL2: Design, compilation and web interface of an online corpus for L2 Spanish acquisition research. *Second Language Research*, *38*(4), 965–983. https://doi.org/10.1177/02676583211050522

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, *15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing*, *34*(4), 493–511. https://doi.org/10.1177/0265532217710675

Mangold, M. (2025). Are my learners advanced? Using measures of complexity and accuracy to analyze L2 Spanish writing. *Foreign Language Annals*, *58*(2), 326–345. https://doi.org/10.1111/flan.12788

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

Menke, M. R., & Strawbridge, T. (2019). The writing of Spanish majors: A longitudinal analysis of syntactic complexity. *Journal of Second Language Writing*, *46*, Article 100665. https://doi.org/10.1016/j.jslw.2019.100665

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, *30*(4), 555–578. https://doi.org/10.1093/applin/amp044

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

Ortega, L. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann, & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127–155). De Gruyter. https://doi.org/10.1515/9783110229226.127

Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, *29*, 82–94. https://doi.org/10.1016/j.jslw.2015.06.008

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*(4), 590–601. https://doi.org/10.1093/applin/amp045

Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117–134. https://doi.org/10.1177/0267658314536435

Paquot, M. (2019). The phraseological dimension in interlanguage complexity research. *Second Language Research*, *35*, 121–145. https://doi.org/10.1177/0267658317694221

Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes*, *14*, 48–59. https://doi.org/10.1016/j.jeap.2013.12.001

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing*, *26*, 10–27. https://doi.org/10.1016/j.jslw.2014.09.003

Qin, W., & Uccelli, P. (2016). Same language, different functions: A cross-genre analysis of Chinese EFL learners' writing performance. *Journal of Second Language Writing*, *33*, 3–17. https://doi.org/10.1016/j.jslw.2016.06.001

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Restrepo-Ramos, F. (2021). Writing to increase complexity: Spanish L2 learners' complexity gains in a college composition class. *Hispania*, *104*(4), 671–690. https://doi.org/10.1353/hpn.2021.0120

Ryshina-Pankova, M. (2015). A meaning-based approach to the study of complexity in L2 writing: The case of grammatical metaphor. *Journal of Second Language Writing*, *29*, 51–63. https://doi.org/10.1016/j.jslw.2015.06.005

Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In P. Bański, H. Biber, E. Breiteneder, H. Kupietz, H. Lüngen, & A. Witt (Eds.), *Proceedings of Challenges in the Management of Large Corpora 3 (CMLC-3)* (pp. 28–34). Association for Computational Linguistics. http://rolandschaefer.net/?p=749

Schnur, E., & Rubio, F. (2021). Lexical complexity, writing proficiency, and task effects in Spanish dual language immersion. *Language Learning & Technology*, *25*(1), 53–72. https://doi.org/10.64152/10125/73425

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. https://doi.org/10.1093/applin/amp047

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics*. Pearson Education.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Second Language Teaching & Curriculum Center, University of Hawai'i at Manoa.

Yang, W., Lu, X., & Weigle, S. C. (2015). Different topics, different discourse: Relationships among writing topic, measures of syntactic complexity, and judgments of writing quality. *Journal of Second Language Writing*, *28*, 53–67. https://doi.org/10.1016/j.jslw.2015.02.002

Yoon, H.-J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, *66*, 130–141. https://doi.org/10.1016/j.system.2017.03.007

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing*, *47*, 100505. https://doi.org/10.1016/j.asw.2020.100505

Zhang, X., & Lu, X. (2022). Revisiting the predictive power of traditional vs. fine-grained syntactic complexity indices for L2 writing quality: The case of two genres. *Assessing Writing*, *51*, Article 100597. https://doi.org/10.1016/j.asw.2021.100597

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426

Pengzhan Yang is undertaking his Ph.D. in Education at the University of Sydney. He has a particular interest in applied linguistics, English-medium instruction, Computer-assisted language learning (CALL), and AI-assisted English language learning. He has published research articles in Social Sciences Citation Index-indexed journals, such as *British Journal of Educational Technology, ReCALL, Teaching and Teacher Education, Journal of Second Language Writing*, and *Journal of English for Academic Purposes*.

Wenqian Huang is a Ph.D. candidate in the Faculty of Arts and Social Sciences at the University of Sydney. She has a particular interest in computer-assisted language teaching and learning, willingness to communicate, Teaching English to Speakers of Other Languages (TESOL), and second language acquisition. She has published research articles in SSCI-indexed journals, such as *British Journal of Educational Technology, ReCALL, Teaching and Teacher Education, Journal of Second Language Writing*, and *Journal of English for Academic Purposes*.

Guangyuan Yao is an Associate Professor in the School of Foreign Languages at Central South University. His research focuses extensively on the applications of artificial intelligence in linguistics and translation. He has published more than 20 articles in leading international journals, including *Applied Linguistics, Journal of Second Language Writing, Assessing Writing, Innovation in Language Learning and Teaching, Journal of Pragmatics*, and *Journal of English for Academic Purposes*, among others.

Cite this article: Yang, P., Huang, W., & Yao, G. (2026). Linguistic complexity and writing proficiency in assessing L2 Spanish writing. *Language Teaching*, 1–17. https://doi.org/10.1017/S0261444826101426

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 07:38:53, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0261444826101426