



## Large and Fine-Grained Complexity Measures and Writing Quality among Iranian EFL Learners

Rajab Esfandiari<sup>1\*</sup>, Mohammad Ahmadi<sup>2</sup>

<sup>1\*</sup>Associate Professor, Department of English Language Teaching, Imam Khomeini International University, Qazvin, Iran

<sup>2</sup>Assistant Professor, Department of English Language, Lorestan University, Khorramabad, Iran

Received: July 31, 2023

Accepted: August 20, 2024

### Abstract

In recent years, there has been extensive research on syntactic complexity in L2 writing. However, a significant research gap remains in understanding the specific association between syntactic complexity measures and the quality of writing among adolescent learners of English as a foreign language (EFL). This research study aimed to bridge this gap by examining the connection between syntactic complexity measures and writing quality in 146 argumentative essays by Iranian-speaking learners of English as a foreign language, as assessed by human raters. Syntactic complexity measures were analyzed through computer programs called L2SCA and TAASSC. Two regression analyses were run to assess the relationship between the predictor and outcome variables. The results revealed that syntactic complexity model explained 35.6% of the variance in the outcome variable (students' writing scores) with "dependent clauses per T-unit", "Mean length of T-unit", "Mean length of clause", and "complex nominals per T-unit" as the strongest predictors. Also, the combined model accounted for 44.1% of the variance in the outcome variable. The findings from this research offer valuable insights into writing assessment and pedagogy, enabling language educators to design more effective instructional strategies tailored to the specific needs of this learner group.

**Keywords:** Complexity Measures; Quality of Writing; Sophistication; Development

### INTRODUCTION

Syntactic complexity (SC) refers to an individual's ability to strategically manipulate morphosyntactic resources to effectively convey meaning (Bulté & Housen, 2012). In second language acquisition (SLA) research, SC has been conceptualized through the complexity theory of SLA (Ortega, 2003). This theory proposes that language development involve nonlinear growth of complexity throughout learners' interlanguage systems as proficiency increases over time.

Advanced complexity emerges through restructuring of underlying grammar rather

than simple linear accumulation of structures (Ellis & Larsen-Freeman, 2009). According to this view, the variety and sophistication of syntactic devices that learners employ offer insights into developing the command of the target language (Housen et al., 2005). However, the examinations of SC through this theoretical lens have often relied on coarse-grained measures failing to capture its multidimensionality (Housen & Kuiken, 2009).

Dynamic Systems Theory (DST) offers a complementary framework for understanding SC development as a complex, dynamic process (Larsen-Freeman & Cameron, 2008). Key tenets of this theoretical model posit that language acquisition arises from intricate interplay

\*Corresponding Author's Email:  
[esfandiari@hum.ikiu.ac.ir](mailto:esfandiari@hum.ikiu.ac.ir)

between countless cognitive, social, and contextual variables within the learning system (Herdina & Jessner, 2002). Subtle differences in initial conditions can cascade into significant changes as proficiency nonlinearly emerges over time (Larsen-Freeman, 2006). Finally, short-term fluctuations reveal developmental patterns characterized by stability/flux phases (Bulté & Housen, 2014; Verspoor et al., 2008).

By observing learners at fine-grained intervals (see Li et al., 2023), interconnected modifications across linguistic subsystems like grammar, lexis and discourse become apparent (de Bot et al., 2008; Larsen-Freeman, 2006). This perspective provides a more nuanced view of emergent, variable gains contradicting rigid stage models of proficiency (de Bot, 2007). As a theoretical driver that shapes SLA research agendas, DST principles have generated interest in microgenetic techniques (Herdina & Jessner, 2002).

An abundance of research has examined the relationships between SC and constructs such as overall language ability (Ortega, 2003), writing quality (Crossley & McNamara, 2014), nativeness (Susoy, 2023), and genre/disciplinary variations (Biber, 2006). However, studies have often failed to scrutinize complexity through the integrated lenses of relevant theories (Housen et al., 2012).

Particularly, investigations dominated by isolated measures like clausal indices, which are often called large-grained measures, as they have limitations in capturing dimensions of complexity beyond clausal features, overlook complexity's multidimensionality (Kyle, 2016; Wolfe-Quintero et al., 1998). While necessary, broad grammatical measures alone cannot detect localized progress influencing proficiency judgments (Bulté & Housen, 2014; Lu, 2011). Finer-grained variables (such as the average number of modifiers per head noun, the average number of dependents per nominals, or the standard deviation of dependents per nominal subjects; See Kyle & Crossley, 2018) address this limitation, yet minimal research adopts both granularities guided by overarching theoretical frameworks (Römer, 2009).

The relationships between SC development and important constructs like writing quality

have received little cross-linguistic focus (Martínez, 2018). Accordingly, there is a need to clarify whether syntactic development contributes to better writing, as reflected in higher-rated papers in terms of text quality, which refers to the overall assessment of a written text, taking into account coherence, organization, clarity, grammatical accuracy, and adherence to genre-specific conventions (Yang & Weigle, 2015). What is more, adolescent EFL contexts remain underexplored despite instructional relevance. A deeper understanding of interactions shaping proficiency paths in this population can empower targeted pedagogical support (Polat, 2013).

Additionally, limited research has employed theoretical models integrating multidimensional SC analyses, or Dynamic Systems methodologies, revealing subtle developmental patterns (Al-Hoorie, et al., 2023). Addressing these gaps holds promise for enhancing conceptualizations of the intricate, dynamic processes underlying SLA.

Considering the existing gap in previous studies, the present study investigates the relationship between syntactic complexity and writing quality in adolescent English as a foreign language (EFL) writing. Writing quality is broadly defined as the inclusion of “clear arguments, ideas that are thoroughly explained, interesting vocabulary, and varied sentence structure” (Nobles & Paganucci, p. 21) in a particular text. The study, accordingly, incorporates the complexity theory of second language development (Ortega, 2003) as a theoretical perspective, which suggests that language development involves the growth and restructuring of complexity in learners' linguistic systems. The ability to produce more syntactically complex structures is seen as an indicator of developing proficiency. By adopting a multifaceted approach that considers different dimensions and subdomains of syntactic complexity, the study aims to comprehensively analyze how syntactic complexity manifests in adolescent EFL writing and its impact on writing quality. Additionally, the study integrates Dynamic Systems Theory (DST) (See Dong, 2016) to uncover nuanced features of language learning systems and understand how learners'

language competence evolves through complex interactions with other variables. This theoretical framework enables a comprehensive analysis of fine-grained patterns of change in syntactic complexity and their influence on writing quality in adolescent EFL writing. The findings enhance our comprehension of writing proficiency development in adolescent EFL learners and inform pedagogical practices for enhancing their writing skills.

## LITERATURE REVIEW

The significance of syntactic complexity (SC) in the development of both first language (L1) and second language (L2) is now widely acknowledged.

### Conceptualization of SC and its Multidimensionality

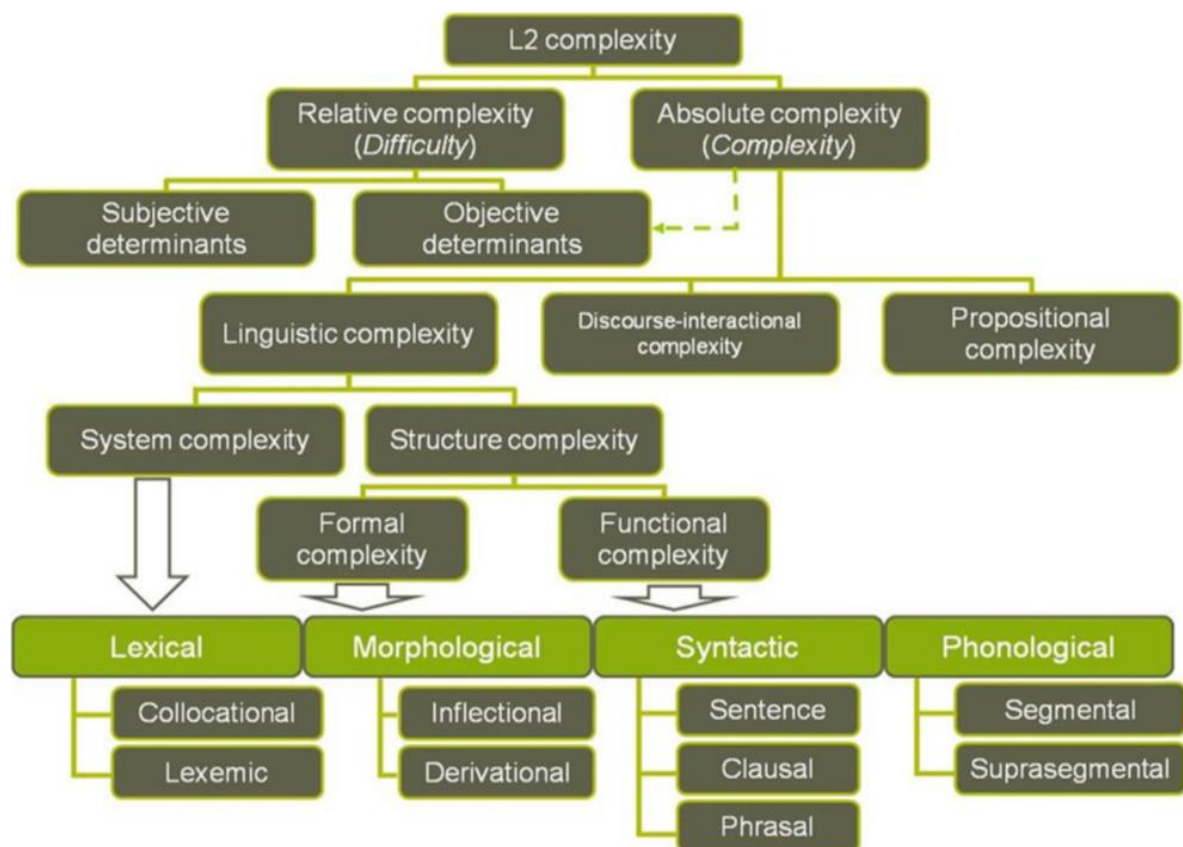
DST presents an alternative perspective for conceptualizing the development of second language complexity (SC) over time. DST views language as a complex, dynamic system wherein multiple interrelated factors interact and influence each other continuously (Larsen-Freeman & Cameron, 2008). A core principle of DST is that small, gradual changes in initial conditions can lead to significant variability in the system as it evolves nonlinearly. When applied to SLA, DST proposes that an individual's developing interlanguage exists within a dynamic system containing cognitive, social, and contextual elements that mutually influence change over time (Larsen-Freeman, 2006).

Longitudinal studies analyzing learner language data through the lens of DST have produced insightful findings. For example, research has shown SC development follows non-linear trajectories characterized by periods of stability and phases of abrupt change (Baba & Nitta, 2014; Verspoor et al., 2008). Analyses of within-learner variability across time have also revealed complex developmental patterns, fluctuations, and interconnected changes between

sub-systems like vocabulary, grammatical structures, and discourse skills (De Bot et al., 2005).

There have been numerous definitions and interpretations of the concept of language complexity. Lennon (1990) defined complexity as the tendency to move beyond conventional language forms and to try out new forms in spite of the fact that they may be erroneous. Skehan (1998), on the other hand, held that complexity is realized in a variety of language forms both in underlying and sophisticated structures. Similarly, Wolfe-Quintero et al., (1998) used the notions of variety and elaborateness in performing the task for defining language complexity. While there is a multitude of complexity definitions in the literature, it is primarily concerned with the more developed grammatical features that language learners exhibit as they progress in proficiency (Biber et al., 2011). Kyle and Crossley (2018) commented that “at the syntactic complexity level, complexity has historically been operationalized through large-grained indices that measure complexity at the clause or sentence level (e.g., the length of clauses, T-units, and/or sentences)” (p. 1).

However, “the understanding of the development of SC in second language acquisition is limited by the ability to measure complexity reliably and accurately” (Connor-Linton & Amoroso, 2014, p. 23). Ortega (2003) state that complexity is a complicated component which includes different aspects, dimensions and sub-constructs, each of which needs to be explored in its own sense. The construct validity of complexity is dependent upon its multi-layered nature in a way that “any inaccurate characterization of the term complexity may question the results of the studies” (Norris & Ortega, 2009, p.563). To tackle the multifaceted nature of language complexity, Bulte´ and Housen (2012) proposed a working model (Figure 1) which consists of different parts and approaches to language complexity, which can be used for making decisions on the related studies.



**Figure 1**  
*A Taxonomy of Complexity Constructs*

By capturing "moment-to-moment fluctuations" over short intervals, a dynamic perspective may provide a more nuanced understanding of how interacting factors drive emergent changes in a learner's developing interlanguage complexity over time (Herdina & Jessner, 2002). Going forward, applying DST concepts and analytical techniques from other fields has the potential to advance both theorizing and empirical research on SC development. For example, conducting microgenetic studies capturing data bi-weekly can explore nonlinear trajectories, critical transitional phases, and dynamic correlations between complexity and related variables (De Bot, 2008; Herdina & Jessner, 2002; Verspoor et al., 2008). This may generate novel insights into the intricate, emergent process of gauging and optimizing learners' developing interlanguage complexity.

### Measures of Complexity in L2

According to Lu (2009), metrics for measuring syntactic complexity have become essential research tools in various language-related fields. More specifically, in second language acquisition context, SC has been used to explore

and categorize the learners' proficiency levels (e.g., Casanave, 1994). The main rationale behind using SC for measuring L2 learners' writing proficiency is that as L2 learners progress through more advanced levels of language production, their SC increases (Larsen-Freeman, 1978). Although SC is not the only predictor of writing proficiency (Lu, 2011), it may be one of the most prominent ones, since measures like embedding and subordination encourage the generation of complex ideas (Beers & Nagy, 2009).

Most of the attempts to operationalize SC have centered around T-unit measures (e.g., mean length of T-unit, clauses per T-unit, dependent clause per T-unit, and so on.). However, the criticism against appropriateness of the T-unit metrics for measuring learners' writing SC has been around as early as 1990s when Bardovi-Harlig (1992) undermined the validity of T-unit indices for writing complexity assessment of proficient L2 learners. That is because T-unit metrics could not adequately capture "the knowledge of the learner" (Bardovi-Harlig, 1992 p.391), since they do not encompass syntactic sophistication, conjuncts and



coordinate sentences; rather, they give the learners' writing "too much credit by breaking up sentences" (p. 392). Biber et al. (2011) stated that the commonly held notion that more intricate grammatical structures and the use of embedded clauses lead to greater grammatical complexity, as measured by T-unit indices, does not have sufficient empirical evidence to support it; rather, it is justified intuitively. For students EFL/ESL students, the T-unit has face validity as it aligns with their language training, which often focuses on grammar and precision. (Casanave, 1994).

One of the scholars who rightly criticized the usefulness of T-unit as the reflective of SC was Rimmer (2006), who eloquently argued that traditional grammar testing was reliant upon intuitional bases and suffered from systematic and corpus-informed evaluation. Rimmer further argued that traditional consideration of unit-based indices as strong determiners of grammatical complexity is not empirically well founded since greater information could be packed into shorter language units. Rimmer also criticized clause-based metrics, like C/TU and DC/C, for neither demonstrating the structures below the clause level nor categorizing between different types of subordinations. Although there was a lack of consensus regarding the appropriateness of above-mentioned measures along the entire developmental trajectory, previous SC studies have frequently used these measures as a proxy for measuring proficiency and development as usually characterized through rated writing quality which will be discussed in the following section.

As alternative to large-grained SC measures, phrasal complexity is assumed to better reflect advanced writing (Biber et al., 2011). Recent corpus studies have revealed that advanced writing is no longer defined by the frequent use of subordinate clauses, which are now more commonly found in conversational language. Instead, advanced writing is characterized by the extensive use of phrasal expressions and prepositional phrases. It is well illustrated by Halliday (2004), who asserts that in the course of writing development clausal complexity decreases. Highlighting the differences between speech and writing, Halliday (2004) further

contends that although complex written register is characterized by simple clauses and sentences, the nominal groups within them may be enormously long and complex.

Biber and Gray (2011) proposed a developmental stage of noun phrase (NP) complexity derived from the findings of a comprehensive analysis of a corpus study comparing speech and writing. These stages are based on the assumption that early academic writing is characterized by complexity features of spoken register (clausal features), which is followed by the gradual use of phrasal complexity features.

Although Biber et al.'s (2011) NP complexity measures are considered to be adequate in describing and predicting language use, and have been extensively studied in the field, they fail to consider functional properties of lexicogrammatical features of interest. Therefore, in addition to broad measures of syntactic complexity, this study also used detailed indices of phrasal complexity suggested by Kyle (2016). "These measures calculate the average number of dependents per phrase type, the occurrence of specific dependent types, and the average occurrence of specific dependent types in particular types of noun phrases" (Kyle (2016, p 25).

### **SC and Writing Quality in L2 Writing**

Recently there has been a growing trend in L2 research towards the relationship between syntactic development and writing quality as assessed through rated writing essays based on writing rubrics. "In L2 writing Learners' writing quality is characterized through their mastery of complex and sophisticated language features" which is accompanied by appropriate and varied vocabulary control (Ravid & Tolchinsky, 2002, p.14).

Prior research on the relationship between SC and writing quality have yielded inconclusive results. Drawing on a computational tool for measuring SC (Coh-Metrix), Crossley and McNamara examined the relationship between writing development and writing quality. They collected essays from 52 L2 learners in a longitudinal study. It was found that growth in SC was the function of time studying English. However, only one of the SC measures that

were used to show writing development was indicative of writing quality. Based on the interpretation of their findings they noted that human raters' assessment of the writing quality was independent of the measures which are commonly used to gauge writing proficiency; rather, they tend to be based on the metrics aligned with spoken discourse. Conversely, Yang et al. (2015) reported that coordination-based measures and particular nominal structures are able to predict writing quality. Casal and Lee (2019) showed that Mean length of T-unit, mean length of clause and complex nominals per clause are good indicators of L2 writing quality. Staples and Reppen (2016) examined the relationship between the language used in first-year writing across three L1s and language ratings using lexico-grammatical approach which identified the vocabulary and grammar that student writers used. The essays were rated by experienced writing teachers for language use. The results indicated that there were important similarities in the use of lexico-grammatical features among the writers of three L1 groups.

### **Purpose of the Study**

Investigations into SC in writing, in general, and its association with writing quality in second language learners, in particular, can potentially improve our comprehension of the linguistic traits that contribute to good writing. However, these studies have limitations from both theoretical and methodological perspectives. (Bi & Jiang, 2020). Ortega (2012) noted that in most of previous studies, SC complexity was investigated as dependent variable and the learners' proficiency levels were independent variable. Put it differently, the bulk of previous research converted the interval variables (learners' proficiency scores) into categorical ones (i.e., the learners' proficiency levels). As a result, statistical procedures such as ANOVA were employed to find statistical significance among the groups. Ortega also observed that this practice had always been criticized by statisticians because of some problems such as "likelihood of Type II error and the lack of power. Thus, some more advanced statistical techniques, such as regression are called for" (Bi & Jiang, 2020, p.35). As a result, and in response to the

call of Plonsky and Oswald (2017) "to use regression as an alternative to ANOVA" (p. 583), the present study intends to fill this void and is designed to explore SC of the texts produced by adolescent EFL learners in relation to writing quality as assessed by human raters. Accordingly, the following research questions are formulated.

***RQ1:** To what degree could large-grained SC metrics predict the writing quality of adolescent intermediate EFL learners?*

***RQ2:** To what degree could incorporating measures of fine-grained SC enhance the predictive capacity of SC measures in assessing the writing proficiency of adolescent EFL learners?*

## **METHODOLOGY**

### **Participants**

The age range for the participating students was between 12 and 19 ( $M = 16.23$ ). They were all of intermediate English proficiency (B1 or B1+). The participants' proficiency levels were determined based on the scores they obtained in an Oxford Placement Test prior to the study. The participants were selected using a convenience sampling method, considering the institutes' willingness to participate and their accessibility to the researchers. The students wrote an argumentative essay on the topic "Nowadays, social media has become an integral part of people's lives. However, there is a lack of consensus regarding its overall impact and effectiveness. Do you believe that social media do more harm than good?"

### **Instruments and Materials**

The instruments consisted of holistic ratings for evaluating writing quality and an automatic computer program called L2 Syntactic Complexity Analyzer (L2 SCA) for analyzing syntactic complexity. The data in this study consisted of 146 argumentative essays produced by Iranian-speaking EFL learners at six private language institutes in Karaj, Iran. The students were given 30 minutes to complete their writing task and were required to write at least 250-300 words on the topic described in the previous part. This topic was selected because it was the

most preferred topic among the students. By controlling the variables of time and topic, the results were made comparable. The essays were distributed randomly in order for the essays to be rated blindly by the independent raters of the study. Prior to the rating process, the raters were trained by the authors of the current research to rate the essays reliably.

### Measures

The essays were measured in two separate phases. In phase 1, The essays were assessed using holistic ratings of writing quality. For holistic rating of the essays, we followed Beers and Naggy's (2009) criteria for assessing writing which included "a) the writer's focus on the topic, (b) supporting details/elaborations, (c) uses of effective language and word choice, and (d) the writer's tone or voice, rated on a scale of 0-5" (p. 192). We decided to choose holistic rating as it is "the preferred and habitual choice in these teachers' daily practice, in part because it makes assessing and giving feedback to work submitted by large classes of students more manageable than if a detailed rubric of marking criteria were applied" (Hoang & Boers, 2018, p. 3). Analytical rating was not the focus of the present study because we analyzed how L2 writers' complexification of the texts and the readers' general impression of writing quality are associated. The compositions were rated by three experienced raters with an MA degree in linguistics.

To assess the syntactic complexity of the English texts produced by the EFL learners in this study, the researchers selected five indices from Lu's (2010) classification of 14 measures. These five indices were chosen in a way that the multidimensionality of the construct of SC could be reliably captured as suggested by Norris and Ortega (2009) (see Table 1).

Due to the importance of phrasal complexity in written register (See Biber et al. 2011), we also assessed the texts in terms of fine-grained phrasal measures of SC as suggested by Kyle (2016). These measures were assessed through TAASSC which is available for free for research and educational purposes. "It takes plain text files as input and produces a comma-separated values (.csv) file as output" (P. 34). Overall,

TAASSC is a comprehensive tool that can be utilized to analyze the syntactic complexity of student writing.

### Data Analysis

As the first step in the analysis of the essays, 10 percent of the texts was rated by the raters independently. Afterwards, the inter-rater agreement was calculated at 0.812. After resolving uncertain cases, the inter-rater reliability rate was almost 87%, which is considered high based on inter-rater agreement reports from other studies (such as Covington et al., 2006). Any remaining discrepancies were discussed until a consensus was reached. The two raters were relatively reliable regarding the ratings of writing quality as measured by intra-rater reliability coefficient which ranged from a low of 0.831 to a high of 0.941.

In the second phase of evaluating the students' essays, we used an automatic computer program called L2 Syntactic Complexity Analyzer (L2 SCA) (Lu, 2010). L2 SCA is a text analyzer tool which is run on Python environment and can compute 14 measures of SC. As Lu (2010) stated, L2CA integrates Stanford Parser (Klein & Manning 2003), which is a "state-of-the-art syntactic parser due to its remarkable accuracy, ease of use, and its built-in sentence segmentation, tokenization, and POS tagging, and Tregex to query syntactically-parsed language samples" (Levy & Andrew, 2006, p. 484). Lu further pointed out that L2SCA was highly reliable in locating the syntactic structures which was reported to be between 0.830 and 1.000. The correlation between the scores for syntactic complexity calculated by the system and those assigned by the human annotators was also exceptionally high, ranging from 0.834 to 1.000.

As for the second research question, two regression analyses were carried out. As discussed earlier, regular inferential statistics (e.g., ANOVA, MANOVA, etc.) tend to ignore a great deal of information which may result in some problems such as Type II error and the power of the test (Ortega, 2012). As alternative to regular inferential statistics, in this study regression analysis is used to predict syntactic complexity. With regard to the first research

question, we built on five SC measures as predictor variables to predict writing scores as assigned by human raters. As for the second research question, all SC indices (large-grained traditional and fine-grained phrasal) were combined in a single model to predict writing quality. As there was a single predictor variable in research question one, a simple regression analysis was employed. However, for the second model (the combined model) a stepwise multiple regression analysis was used. Before conducting the main analysis, preliminary assumption testing was performed to check for normality, linearity, and multicollinearity. The results of the testing indicated that there were no significant violations of these assumptions.

## RESULTS

Table 1 below summarizes the descriptive statistics of five SC measures. The students' essays had a mean length of 7.34 words per clause, 10.98 words per T-unit, and 11.90 words per sentence. In addition, there were 1.24 clauses per sentences, 0.51 dependent clauses per T-unit, and 1.08 complex nominals per T-unit.

In traditional SC model four measures of DC/T, MLT, MLC, and CN/T ( $R^2 = 0.356$ ,  $F_{142} = 57.32$ ,  $p = .001$ ) made statistically significant contribution to the prediction of dependent variable. As a result, 35.6% variance in the writing quality of the students could be accounted for by 4 SC measures. As shown in Table 2, the coefficient of the four predictors in this model were, 1.412, 0.313, 1.112, and 1.347 for DC/T, MLT, MLC, and CN/T respectively. This means that one unit of increase in the dependent clauses in each T-unit, number of words in each T-unit, number of words in each clause, and number of complex nominals in each T-unit will increase the writing quality of the students' essays by 1.412, 0.313, 1.112, and 1.347 points respectively. This suggests that the greater syntactic complexity values, the higher the writing quality.

For the second research question, we chose a number of fine-grained measures that are used in TAASSC. Having checked the preliminary assumptions, we finally came up with 12 measures out which 2 variable made statistically unique contribution to the prediction of

our final model. Accordingly, the final model included 4 large-grained traditional SC measures (DC/T, MLT, MLC, and CN/T) and 2 fine-grained phrasal complexity measures (dependents per nominal, dependents per nominal subjects). Totally, the variable explained 41.7% of the variance in writing quality of the students' essays ( $R^2 = 0.441$ ,  $F_{142} = 78.56$ ,  $p = .000$ ). Four predictors of DC/T, MLC, dependents per nominals, dependents per nominal subject remained in the resulting model. According to Table 3 an addition of DC/T, dependents per nominals, dependents per nominal subject, and words per clause units, would cause the writing scores to increase by 1.412, 1.126, 0.381, and 0.275 points respectively. To sum up, in the combined model, the strongest predictor variable was DC/T which was a strong predictor of the students' writing scores. The comparison of traditional SC model and the combined model based on  $R^2$  values revealed that the combined model outperformed the traditional SC model, explaining 8.8% more variance in students' writing quality scores. However, the results of Fisher  $r$  to  $z$  transformation indicated that this difference was not significant suggesting that the addition of fine-grained phrasal complexity measures increases the predictive power of power of SC measures.

**Table 1**  
*Descriptive Statistics for Five SC Measures*

Measures	Mean	SD
MLT	10.98	2.11
MLC	7.34	1.02
C/S	1.24	0.27
DC/T	0.51	0.18
CN/T	1.08	0.12

**Table 2**  
*Regression Coefficients in the Traditional SC Model*

Predictors	$b(\beta)$	$p$
DC/T	1.412	0.000
MLT	0.313	0.000
MLC	1.112	0.012
CN/T	1.347	0.031



**Table 3**  
**Regression Coefficients in the Combined Model**

Predictors	b( $\beta$ )	p
DC/T	1.412	0.000
Dependents per nominals	1.126	0.000
Dependents per nominal subjects	0.381	0.012
MLC	0.275	0.031

## DISCUSSION

In response to the first research question, which explored the predictability of large-grained syntactic complexity (SC) measures in relation to writing quality, the study found that four measures: DC/T, MLT, MLC, and CN/T, significantly contributed to the prediction of the dependent variable (writing scores assigned by human raters). These measures have been widely used in previous research to capture different aspects of syntactic complexity. Norris and Ortega (2009) argued that MLT measures SC at the global level, reflecting overall complexity, while MLC measures complexification at the phrasal level. Global syntactic complexity measures such as MLT, however, do not “specify which type of modification contributes to the complexity of the texts” (Kreyer & Schaub, 2018, p. 93). Therefore, they are sometimes referred to as “omnibus measures as they combine complexity at phrasal and clausal levels” (Biber et al., 2020, p. 12).

The finding should also be explained in terms of the theoretical background to the study. Previous studies on the relationship between MLT, DC/T, and writing proficiency have yielded mixed results. Some studies have shown a strong association between the length of T-units (DC/T) and writing quality, while others have found no or weak relationship. This inconsistency can be partially explained by the dynamic system theory (DST) to learning (See Lei et al., 2023), which posits that language development is not a discrete stage-like process but rather a dynamic and adaptive process with waxing and waning patterns (Larsen-Freeman, 2006). Because of the dynamic nature of learners' syntactic repertoires at different acquisition trajectories, language learners may complexify texts using different syntactic devices. This

suggests that “the results of previous studies targeting advanced college-level EFL learners may not be applicable to less proficient EFL learners” (Bi & Jiang, 2020, p. 37).

The contrasting findings in previous studies could also be attributed to differing operationalizations of variables, variability in research design, learners' first language (L1), and language proficiency, which lead to inconsistent results on particular measures of SC. Wolfe-Quintero et al. (1998), for example, argued that when definitions differ, the constructs being measured are also different, making it challenging to cross-compare research findings. In the present study, different statistical metrics were used, with some studies employing conventional statistical tests such as ANOVA, while others used more powerful statistical tools such as regression analysis. These methodological differences can contribute to variations in results. Furthermore, the influence of learners' L1 has been found to affect the use of specific syntactic structures, as different L1 backgrounds may lead to the underuse or overuse of certain syntactic features (Bi & Jiang, 2020).

MLC captures complexity at the subclausal level through phrasal elaboration (Norris & Ortega, 2009). The underlying idea is that “increases in clause length reflect increases in phrase length through modification of the head” (Bulté & Housen, 2012, p. 34). The finding that MLC and the specific phrasal complexity measure (CN/T) were significant predictors of writing quality is consistent with earlier research on L2 complexity. Bulté and Housen (2014), Lu (2011), and Yang et al. (2015) have also found that phrasal complexity measures are indicative of advanced writing proficiency. According to Biber et al. (2011), progression in writing moves from clausal to phrasal elaboration. Therefore, the inclusion of phrasal complexity measures as predictors of writing quality aligns with theoretical expectations.

To address the second research question, we added fine-grained phrasal complexity measures to the combined model and the results revealed that 8.8% more variance in the writing quality of the students' essays was explained, although it was not statistically significant. These findings underline the importance fine-grained

phrasal complexity measures as important indicators of EFL learners writing proficiency. The results of the analysis of fine-grained phrasal complexity measures suggest that more advanced writers tend to use more dependents in their essays, which is consistent with Biber et al.'s (2011) argument that more noun-modifying features characterize advanced writing. These findings are also supported by the results of Crossley and McNamara's (2014) study, which found that "over the course of a semester, L2 learners produced texts that aligned with the features of academic writing, including more nouns and phrasal complexity" (p. 76). However, it should be noted that human raters evaluated the texts solely based on the features indicative of spoken register, such as clausal complexity. The number of dependents per nominal subject in the present study also contributed to the complexity of the text. This implies that "in order for student writers to appear more advanced in the eyes of evaluators and human raters, they need to include fewer bare nominal subjects and more structures with multiple dependents in the subjective position" (Esfandiari & Ahmadi, p. 14).

However, in this study, clausal complexity measures showed a stronger predictive power than phrasal complexity measures. This finding suggests that while human raters are aware of the importance of phrasal features in advanced writing, they still tend to prioritize clausal features, which are more commonly observed in conversation than in writing. This finding is consistent with that of Crossley and McNamara (2014), who reported that despite significant advancements in phrasal complexity, clausal complexity remains a strong predictor of writing quality.

The results of the present study support the idea that a combination of both clausal and phrasal complexity measures is necessary to capture the multidimensional nature of syntactic complexity. The use of multiple measures allows for a more comprehensive understanding of learners' syntactic development and its relationship to writing quality. This finding confirms some of those of previous research that emphasizes the importance of considering multiple dimensions of syntactic complexity to gain a more

accurate picture of learners' proficiency (Bulté & Housen, 2012; Crossley & McNamara, 2014).

The theoretical background of the study offers further insights into the findings. The dynamic system theory (DST) approach to learning highlights the dynamic and adaptive nature of language development, suggesting that learners may employ different syntactic devices at different stages of proficiency. This aligns with the mixed results found in previous studies as the syntactic repertoires of learners vary based on their acquisition trajectories (Larsen-Freeman, 2006; Bi & Jiang, 2020; Zhang & Lu, 2022). The primary focus of the study was on intermediate-level EFL learners, whose syntactic development may differ from advanced college-level learners. Therefore, it is crucial to consider the specific learner population when interpreting the results and comparing them with previous studies.

Moreover, the influence of learners' first language (L1) on syntactic complexity should not be overlooked. Learners from different L1 backgrounds may exhibit different patterns of syntactic usage due to the transfer of structures or the influence of L1 syntactic features. This variability in L1 backgrounds can contribute to variations in the results of studies examining syntactic complexity measures (Bi & Jiang, 2020). Future research could explore the impact of L1 on syntactic complexity in more detail to gain a comprehensive understanding of the interplay between L1 and L2 syntactic development.

## CONCLUSION

The present study examined 146 argumentative essays produced by intermediate adolescent EFL learners in terms of writing quality which was operationalized by the scores assigned by human raters. We followed Norris & Ortega's (2009) methodology and employed multiple metrics of syntactic complexity, with each metric intended to measure a specific facet of the overall construct of syntactic complexity. The findings revealed that clausal, global, and phrasal measures (DC/T, MLT, MLC, and CN/T) had statistically significant unique contribution to the prediction of writing quality, explaining 35.6% variance in the dependent variable. The

results obtained from the second phase of the study showed that fine-grained measures of phrasal complexity increased the power of the test by 8.8% although the improvement was not statistically significant.

To summarize, the findings present additional support for the use of different measures to assess EFL learners' writing proficiency (Norris & Ortega, 2009; Ortega, 2003). In other words, SC is not a unidimensional construct, but rather a multidimensional one that should be measured using distinct sources of complexification (Youn, 2014).

The study's results also highlighted the importance of length-based measures of SC in EFL adolescent writing, as longer structures were found to be more complex, as indicated by longer T-units and phrasal structures.

While Biber et al., (2011) argued that development in writing moves away from clausal features to phrasal features, as this study and other studies (e.g., Crossley & McNamara, 2014; Esfandiari & Ahmadi, 2021) showed, we need to be cautious about overemphasis on phrasal features at the expense of clausal features. This study found clausal measures had stronger predictive power than phrasal features for intermediate proficiency. This aligns with the dynamic systems view that syntactic development follows non-linear, adaptive patterns rather than strict stages (Larsen-Freeman, 2006). Features influencing texts may vary depending on the proficiency level.

In the present study, clausal complexity indices offered "more progress-sensitive account of academic writing quality than phrasal measures" (Esfandiari & Ahmadi, p.40). In addition, owing to the fact various developmental factors may interact leading to nonlinear U-shaped trends of some measures (Bulté & Housen, 2012), any single measure may not adequately represent the development of construct of SC at different stages of writing development. Accordingly, global measures and specific measures (length-based, clausal and phrasal) may well complement each other for assessing adolescent EFL learners' writing proficiency (See Kuiken & Vedder, 2012).

According to Casal and Lee (2019), this study proposes that advanced academic writing

is enhanced by a wider range of complex structures and improved functional abilities within those structures. The study also suggests that the predictability of complex language features, specifically measures of complexity based on length and clauses, indicates that providing explicit instruction on these features is likely to assist intermediate English as a Foreign Language (EFL) learners in developing a more intricate writing style that is considered advanced by human evaluators.

In order to accomplish this, instructional tasks that raise students' awareness can be created to highlight the importance of noun phrase modifiers in academic writing and their specific syntactic functions, including nominal subjects and direct objects (as suggested by Lu & Wu, 2022). Moreover, students' writing can be assessed by comparing it to that of advanced writers, focusing on key syntactic features like the average length of independent clauses (T-units) and the number of dependents per nominal. Any loosely-structured syntactic constructions can be collectively adjusted through collaborative efforts to meet the standards of advanced writing.

While this study offers novel insights into a multidimensional complexity assessment, certain limitations point to avenues for future research. The modest corpus size precluded in-depth analysis at the individual learner level or across writing topics/genres. Broader data sets would allow exploring intricacies of how proficiency sub-levels and first language backgrounds influence syntactic choices. Investigating lexical and discourse-level constructs in tandem could provide a more comprehensive view of developing writing ability.

Longitudinal or classroom-based studies may offer deeper understanding of developmental trajectories by closely observing syntactic growth over short intervals. The dynamic nature of language learning suggests periodic evaluation better reflects non-linear progression than isolated pre/post comparisons.

Generally, this study contributes to conceptualizing and measuring syntactic complexity as it relates to proficiency. While not generalizable beyond the research scope, findings seek to advance assessment tools and pedagogical methods supportive of intermediate EFL learners'

continuing development as capable academic writers.

## References

- Ahmadi, M., Esfandiari, R., & Zarei, A. A. (2020). A corpus-based study of noun phrase complexity in applied linguistics research article abstracts in two contexts of publication. *Iranian Journal of English for Academic Purposes*, 9(1), 76-94.
- Al-Hoorie, A. H., Hiver, P., Larsen-Freeman, D., & Lowie, W. (2023). From replication to substantiation: A complexity theory perspective. *Language Teaching*, 56(2), 276-291.
- Bardovi-Harlig, K. (1992). The relationship of form and meaning: A cross-sectional study of tense and aspect in the interlanguage of learners of English as a second language. *Applied Psycholinguistics*, 13(3), 253-278.
- Baba, K., & Nitta, R. (2014). Phase transitions in development of writing fluency from a complex dynamic systems perspective. *Language Learning*, 64(1), 1-35.
- Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing*, 22(2), 185-200.
- Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing*, 24(2), 183-202.
- Bi, P., & Jiang, J. (2020). Syntactic complexity in assessing young adolescent EFL learners' writings: Syntactic elaboration and diversity. *System*, 91, 102248.
- Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge University Press.
- 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.
- Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing*, 26(1), 42-65.
- Bulté, B., & Housen, A. (2012). Defining and operationalizing L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). John Benjamins Publishing Company.
- 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.
- Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, 3(3), 179-201.
- Connor-Linton, J., & Amoroso, L. (Eds.). (2014). *Measured language: Quantitative studies of acquisition, assessment, and variation*. Georgetown University Press.
- Council of Europe. (2001). *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge University Press.
- Covington, M. A., He, C., Brown, C., Naci, L., & Brown, J. (2006). How complex is that sentence? A proposed revision of the Rosenberg and Abbeduto D-Level Scale. In *Caspr research report 2006-01*. Athens, GA: The University of Georgia, Artificial Intelligence Center.
- 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(1), 66-79.
- De Bot, K. (2008). Introduction: Second language development as a dynamic process. *The Modern Language Journal*, 92(2), 166-178.
- De Bot, K., Lowie, W., & Verspoor, M. (2005). *Second language acquisition: An advanced resource book*. Psychology Press.
- Dong, J. (2016). A dynamic systems theory approach to development of listening strategy use and listening performance. *System*, 63, 149-165.



- Ellis, N. C., & Larsen-Freeman, D. (2009). *Language as a complex adaptive system* (Vol. 11). John Wiley & Sons.
- Esfandiari, R., & Ahmadi, M. (2021). Syntactic complexity measures and academic writing proficiency: A corpus-based study of professional and students' prose. *The Journal of AsiaTEFL*, 18(3), 745-763.
- Halliday, M. A. K. (Ed.). (2004). *Lexicology and corpus linguistics*. Continuum.
- Herdina, P., & Jessner, U. (2002). *A Dynamic Model of Multilingualism: Perspectives of Change in Psycholinguistics*. Multilingual Matters.
- Hoang, H., & Boers, F. (2018). Gauging the association of EFL learners' writing proficiency and their use of metaphorical language. *System*, 74, 1-8.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Housen, A., Pierrard, M., & Van Daele, S. (2005). Structure complexity and the efficacy of explicit grammar instruction. *Investigations in instructed second language acquisition*, 12(1), 235-269.
- Ke Li, Y., Lin, S., Liu, Y., & Lu, X. (2023). The predictive powers of fine-grained syntactic complexity indices for letter writing proficiency and their relationship to pragmatic appropriateness. *Assessing Writing*, 56, 100707.
- Klein, D., & Manning, C. D. (2003, July). Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting of the association for computational linguistics* (pp. 423-430).
- Kreyer, R., & Schaub, S. (2018). The development of phrasal complexity in German intermediate learners of English. *International Journal of Learner Corpus Research*, 4(1), 82-111.
- Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143-170). John Benjamins Publishing Company.
- Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic complexity sophistication* [Doctoral dissertation, Georgia State University].
- Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine-grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333-349.
- Larsen-Freeman, D. (1978). An ESL index of development. *TESOL Quarterly*, 12(4), 439-448.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619.
- Freeman, D. L., & Cameron, L. (2008). Research methodology on language development from a complex systems perspective. *The modern language journal*, 92(2), 200-213.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40(3), 387-417.
- Levy, R. & Andrew, G. (2006). Tregex and Tsurgeon: Tools for querying and manipulating tree data structures. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).
- Lei, L., Wen, J., & Yang, X. (2023). A large-scale longitudinal study of syntactic complexity development in EFL writing: A mixed-effects model approach. *Journal of Second Language Writing*, 59, 100962.
- Lu, X. (2009). Automatic measurement of syntactic complexity in child language acquisition. *International Journal of Corpus Linguistics*, 14(1), 3-28.
- 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.
- Lu, X., & Wu, J. (2022). Noun-phrase complexity measures in Chinese and their relationship to L2 Chinese writing quality: A comparison with topic-comment-unit-based measures. *The Modern Language Journal*, 106(1), 267-283.
- Martínez, A. C. L. (2018). Analysis of syntactic complexity in secondary education EFL writers at different proficiency levels. *Assessing Writing*, 35, 1-11.
- 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.
- 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.
- Ortega, L. (2009). Studying writing across EFL contexts: Looking back and moving forward. In R. M. Manchón (Ed.), *Writing in foreign language contexts. Learning, teaching, and research* (pp. 232-255). Multilingual Matters.
- Ortega, L. (2012). Interlanguage complexity. *Linguistic complexity: Second language acquisition, indigenization, contact*, 13, 127.
- Plonsky, L., & Oswald, F. L. (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition*, 39(3), 579-592.
- Polat, B. (2013). Language experience interviews: What can they tell us about individual differences? *System*, 41(1), 70-83.
- Polat, N., Mahalingappa, L., & Mancilla, R. L. (2020). Longitudinal growth trajectories of written syntactic complexity: The case of Turkish learners in an intensive English program. *Applied Linguistics*, 41(5), 688-711.
- Ravid, D., & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language*, 29(2), 417-447.
- Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing*, 23(4), 497-519.
- Römer, U. (2009). The inseparability of lexis and grammar: Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics*, 7(1), 140-162.
- Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics*, 18, 268-286.
- Staples, S., & Reppen, R. (2016). Understanding first-year L2 writing: A lexico-grammatical analysis across L1s, genres, and language ratings. *Journal of Second Language Writing*, 32, 17-35.
- Susoy, Z. (2023). English L1 VS. L2 differences in dissertation abstracts: Lexical density, lexical diversity and academic vocabulary use acuity. *Journal of English Language Pedagogy, Literature and Culture*, 8(2), 17-45.
- Verspoor, M., Lowie, W., & Van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214-231.
- Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. University of Hawaii 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.
- Youn, S. J. (2014). Measuring syntactic complexity in L2 pragmatic production: Investigating relationships among pragmatics, grammar, and proficiency. *System*, 42, 270-287.
- 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, 100597

**Biodata**

**Rajab Esfandiari** is an associate professor of English Language Teaching at Imam Khomeini International University in Qazvin, Iran. His areas of interest, and specialty, include teaching and researching L2 writing, the construction of rating scales, and EAP teaching and testing. He is widely published in international journals, including *Assessing Writing*, *Journal of English for Academic Purposes*, *JALT Journal*, the *JALT CALL Journal*, *Journal of Teacher Education for Sustainability*, *Language Testing in Asia*, the *Journal of Asia TEFL*, and *TESL-EJ*. His most recent paper was published in *Education and Information Technologies*.

Email: [esfandiari@hum.ikiu.ac.ir](mailto:esfandiari@hum.ikiu.ac.ir)

**Mohammad Ahmadi** is a visiting (assistant) professor in the Department of English Language, Faculty of Humanities at Lorestan University, in Khorramabad, Iran. His research interest areas include corpus linguistics, second language writing, computational analysis of natural language, and formulaic language.

Email: [ahmadi.m8362@gmail.com](mailto:ahmadi.m8362@gmail.com))