Syntactic and lexical evolution in L2 writing through the lenses of Dynamic Systems Theory

Reza Kalantari
Research Institute for Education, Iran

Javad Gholami
Urmia University, Iran

This longitudinal, descriptive-exploratory case-study examined Iranian EFL learners' writing complexity through the lenses of Dynamic Systems Theory (DST). One hundred and twenty independent essays written by 12 intermediate to advanced female EFL learners in a TOEFL iBT preparation course over six months constituted the 43,478-word learner-generated corpus of this study. L2 Syntactic Complexity Analyzer was employed to analyze the length of production, sentence complexity, subordination, coordination, and particular structures. Moreover, three lexical analysis software programs including Coh-Metrix, Lexical Complexity Analyzer, and VocabProfile were employed to measure lexical density, diversity, and sophistication. The results of repeated measures analysis of variance (ANOVA) indicated significant differences between time and mean scores in five out of 14 syntactic indices. Correlational analyses among syntactic indices revealed positive relationships among the measures of the same sub-dimension of syntactic complexity. Meanwhile, particular structures enjoyed a positive correlation with both coordination and length of production. The analysis of syntactic and lexical relationships revealed that mean length of sentence, mean length of T-unit and mean length of clause closely corresponded with only lexical diversity. However, these syntactic indices revealed no significant correlations with both lexical density and sophistication. The findings suggest that different syntactic and lexical dimensions interactively comprise L2 writing complexity.

Email address for correspondence: rmkalantari@gmail.com

Keywords: Dynamic Systems Theory, lexical complexity, syntactic complexity

Introduction

Second language development (SLD) study has experienced multiple theoretical and empirical changes in recent decades. One of the important theories of SLD research is Dynamic Systems Theory (DST) which has recently been the topic of much interest. According to Larsen-Freeman (1997), this can even be considered a paradigm shift in applied linguistics. The central idea in DST is that language development is dynamic and complex cognitive demand. For many decades, a leading view to language acquisition has considered language as a modular with separate and independent subsystems for sound, meaning, and structure. However, this view is interpreted as a snapshot approach because of taking into account only one aspect of language at a time. Recently, DST has emerged as a complementary perspective to the cross-sectional and linear approach to applied linguistics (Caspi, 2010; Kyle, 2016; Larsen-Freeman, 2006, 2014, 2019; Verspoor et al., 2011). DST can be considered a theory of second language development that indicates how language, as any other complex system, changes over a time scale or context in which a plethora of factors interact in the emergence, development, and assessment of second language. Writing consists of multiple embedded subsystems, including vocabulary, syntax, semantics, pragmatics, and discourse. In order to obtain a complete picture of L2 writing development and assessment, the relationships among these joint variables warrant further analysis. Such a multicausality approach is useful in providing accurate information on the development of second language writing proficiency (Lowie & Verspoor, 2015).

As an integral manifestation of second language, syntactic complexity has been considered a major construct in L2 writing development studies in recent decades (Kyle & Crossley, 2018; Lu, 2010). However, syntactic complexity does not entirely account for writing complexity, and this seems to be particularly true in the case of EFL writers (Ortega, 2003). To further explore writing complexity, this study sought to examine syntactic and lexical complexities longitudinally in order to identify syntactic and lexical relationships.

Literature review

Dynamic Systems Theory

The idea of complex systems and second language development which can be associated with an intra-paradigm shift in Diane Larsen-Freeman's research orientation (Panahi et al., 2024) is blossoming now, as Ellis (2007) rightly predicted: "a DST characterization of L2 acquisition [development] as an emergent process marks the coming of age of SLA research" (p. 23). In spite of being young, DST has attracted the attention of many researchers (e.g., Bulte´ & Housen, 2012; 2018; Verspoor, et al., 2011). The view of SLA highlights the nonlinearity and dynamic nature of system development. In Larsen-Freeman's (2006) words, the development is not discrete and stage-like, but rather more like the waxing and waning of patterns, sometimes changing continuously, sometimes discontinuously, and even chaotically. Based on DST, language, lending itself par excellence in terms of complex systems, consists of a set of interrelated and interconnected components. Changes in the components of one system will have an effect on all other systems. According to Schmid et al. (2011), learners have to make some kind of choice among the different components in order to convey their message. In the same way, Vercellotti (2015) asserts that in the cognitive assumption of SLD, human's limited attentional resources inhibit learners from concentrating on all language variables simultaneously.

DST theory argues that since language is a complex dynamic system, implementation of traditional methods to evaluate language learning development may not provide reliable results. In order to predict how language development takes place, a large amount of information is needed. As Schmid et al. (2011) convincingly pointed out, DST attempts to gauge a full picture of a learner development across all components of their linguistic proficiency.

Complexity in L2 writing measurement

Syntactic and lexical complexities have long been of interest to L2 writing researchers, since the development of a learner's lexical and syntactic repertoire is thought to be a display of their global proficiency (Kyle, 2016; Lu, 2010; Ortega, 2003). Numerous measures have been proposed to measure large and fine-grained indices of complexity

in L2 writing (e.g., Crossley & McNamara, 2014; Kyle & Crossley, 2018; Ortega, 2015; Norris & Ortega 2009; Shardlow et al., 2022; Wolfe-Quintero et al., 1998). However, the most common indices for syntactic development have been various measures of complexity, accuracy, and fluency (CAF) (e.g., Norris & Ortega 2009; Polat & Kim, 2014). According to Lu and Ai (2015), for a long time, the various measures of accuracy, fluency, and complexity have been identified as fundamental for capturing L2 proficiency. These measures were used to study L2 development using cross-sectional or longitudinal approaches and in naturalistic or instructional settings, applying test-retest design at the commencement and at the end of the course, rather than tracking development during the course.

An overview of the literature shows that it still looks very demanding to find robust, precise and accountable instruments to measure syntactic and lexical development. According to Lu and Ai (2015), most L2 writing teachers and researchers could only manually estimate and account for a limited number of lexical and syntactic indices or components. Among L2 writing development studies, T-units, the shortest complete grammatical unit in writing (i.e., an independent clause and any clauses dependent on it) have been widely used as a L2 development measure (e.g., Larsen-Freeman, 2006). A review of 39 articles on L2 writing revealed that the mean length of T-units is the best index to measure syntactic complexity (Wolfe-Quintero et al., 1998). According to Ortega (2003), the mean length of sentence and the mean number of T-units per sentence are recognized as good indicators to examine syntactic complexity.

However, findings of some studies indicate that more proficient language learners may not necessarily use more T-units or clauses. For instance, Rimmer (2006) convincingly states that syntactic development should reveal development in phrasal features including noun post-modifiers. Consequently, as he suggested, the single T-unit analysis cannot be a complete indicator for L2 development. According to Kyle and Crossley (2018), mean length of T-unit only gives a general indication of how elaborated a particular main clause is but says nothing about the type of elaboration included. Biber et al. (2011) also challenge traditional measures of L2 writing development that equate grammatical complexity with clause subordination (e.g., T-units). Analyzing conversation versus academic writing corpora, they demonstrate that conversation uses more clauses while academic writing favors phrasal complexity

(e.g., complex noun phrases). They argue that L2 assessment tools privileging clausal complexity inadvertently adopt conversational norms, misrepresenting development toward academic proficiency. The study proposes register-appropriate complexity metrics: phrasal features for academic writing and clausal features for speech.

The most recent measurement model that corresponds to this recommendation was put forth by Lu (2010) via designing the L2 syntactic complexity analyzer (L2SCA). According to Mancilla et al. (2015), L2SCA is the most updated software to measure the syntactic complexity of writing, which is especially suited for large-scale corpus-based studies of complexity. Several studies have used this freely available text analysis tool to efficiently analyze written syntactic complexity (e.g., Kyle & Crossley 2018; Lu & Ai, 2015).

Syntactic and lexical constructs are two core building blocks of writing, so it would be more meaningful to see how they could interact over time. A review of the literature suggests three possible types of relationships between syntactic and lexical features: supportive, competitive, and conditional (see Verspoor & van Dijk, 2011). In a supportive relationship, both variables develop in tandem, as the advancement of one facilitates the development of the other. In a competitive relationship, the variables develop in an alternating pattern, with growth in one potentially inhibiting the other (ibid).

The present study aimed to explore L2 writing development longitudinally through the lenses of DST. The major incentive for this study was that, to the best of our knowledge, no study has been conducted on the process of writing complexity development of Iranian EFL learners at the university level. Many Iranian post-graduate students attend TOEFL or IELTS preparation courses, one major module of which is concerned with their writing development. The participants in these courses are highly motivated and try their best to meet the stringent requirements of those high-stake proficiency tests. Moreover, most studies in this realm of inquiry have employed a limited number of lexical and syntactic components to account for writing complexity, while utilizing a single analytical approach encompassing both data collection methods and computational tools. Therefore, building upon the research tradition in these studies, investigating writing development of the same learners through a good number of highly efficient and effective computational analyzers and the triangulation of their

findings would probably yield more robust and accountable results. Therefore, the current study is guided by the following research questions:

RQ1. How does the syntactic complexity of EFL learners' L2 writing develop over time?

RQ2. What are the relationships among the sub-components of syntactic complexity (i.e., length of units, sentence complexity, subordination, coordination, particular structure)?

RQ3. What is the relationship between EFL learners' syntactic and lexical development?

Method

Participants

The participants of this study were 12 female learners of English who had a long experience of language learning in high school, private language schools, and university. The sampling technique of this case study was purposeful (as was the case in Bulte´ & Housen, 2018; Caspi, 2010; Salsbury, 2000; Verspoor, et al., 2011). With purposive sampling, we aimed to identify individuals who were likely to have specific experiences. In this case, the participants in this study were intermediate to advanced English learners, determined by their TOEFL iBT total scores and writing module results. Following Dynamic Systems Theory (Verspoor et al., 2011), participants were selected across intermediate to advanced proficiency levels. The rationale for selecting participants across proficiency levels was justified by research showing that proficient L2 writers demonstrate greater lexical and syntactic diversity (Crossley & McNamara, 2014; Ortega, 2003), and exhibit nonlinear developmental patterns in writing skills (Verspoor et al., 2011). Prior research also (e.g., Verspoor et al., 2008; Spoelman & Verspoor, 2010) demonstrates that learners at different proficiency levels exhibit distinct trajectories in complexity, accuracy, and fluency. As a case, Verspoor et al. (2008) showed that L2 learners exhibit greater variability at intermediate proficiency levels (the "U-shaped" development pattern), supporting the need for diverse proficiency sampling.

The procedure of this study was to demonstrate the development of syntactic complexity through at least ten typed essays over the research period, and to analyze individuals' writings in terms of syntax and lexicon. As a result, it was not feasible to manage larger samples. The participants' age range was between 24 and 37 (see Table 1). They all had Azeri L1 background, studying non-English majors at university. They were all postgraduate students, and their language proficiency level ranged between 70 and 89 (Mean score = 75) out of 120 on the TOEFL iBT. Moreover, their scores in the writing module of the TOEFL iBT with one independent and one integrated writing task were found to vary from 16-22 out of 30 (Mean score = 18). They had studied English in the public education system and university for 10-12 years with limited hours of formal instruction. Meanwhile, these participants had attended private language schools for 3-12 years. In the Iranian context, students study English at schools two to four hours every week with locally developed textbooks which emphasize mostly grammar and reading comprehension with little attention to the development of writing and speaking skills. At the university, students undertake general English and English for Specific Purposes (ESP) courses with a heavy load of reading materials. Due to low efficiency of these formal language learning courses held in crowded classes with few resources, a considerable number of primary, secondary, and even tertiary students attend private language schools to improve their language proficiency in a more communicative and systematic manner (see Naghdipour, 2016). The course instructor had a Master's degree in Teaching English as a Foreign Language (TEFL) and was a highly qualified English teacher with over 20 years of teaching experience in both private and public schools. Based on the school manager, the most qualified teachers were assigned to run proficiency exam preparation courses at their institute.

Writing tasks and data collection procedures

The present case study explored writing development from DST perspective. The design was longitudinal and descriptive-exploratory. To explore the process of writing development, 12 EFL learners of English who have enrolled in a TOEFL iBT writing class were invited to take part in this study. Before participating in this study, they took the TOEFL iBT test. Based on their scores, they were classified as intermediate

and advanced English users. They attended the class twice a week for six months and received instruction on TOEFL iBT independent writing tasks.

Table 1. Participants' profile

Name	Age	Major	Years of L2 learning		TOEFL iBT total Score	TOEFL iBT Writing Score
			School and University	Private Language School		
S-I	27	Economics	10	3	71	16
F-A	24	Medicine	10	12	89	22
E-M	37	Dentistry	10	10	87	22
M-S	25	Medicine	10	4	75	18
Y-A	27	Economics	10	4	73	18
H-B	24	Medicine	10	5	70	18
F-D	24	Medicine	10	5	71	17
F-E	24	Medicine	10	4	70	18
T-H	25	Medicine	10	8	72	18
G-A	29	Dentistry	10	3	72	17
N-F	24	Medicine	10	6	77	16
R-A	26	Medicine	10	9	71	18

The classroom instruction followed a step-by-step process-oriented and simulation-based L2 instruction through feeding, leading, showing, and throwing as main process options (McGrath, 1997). To begin with, the learners received instruction on key aspects of paragraph and essay writing such as topic sentence, thesis statement, paragraph unity, coherence, cohesion, logical progression of ideas, supporting one's ideas, and similar issues from the covered materials accompanied by teacher explanation, tips, and exemplification (feeding). Meanwhile, they were exposed to writing samples or templates with pre- or interactively-highlighted features of those model essays (showing). Later, they were engaged in some guided and staged writing practice activities where they received on-the-spot assistance and scaffolding from their teacher, peers, and available resources such as their dictionaries (leading). Finally, as an integral component of the course, the learners were asked to compose a typical five-paragraph essay including introduction, main/body, and conclusion paragraph in the class on their own (throwing). To simulate the real TOEFL iBT exam conditions as was the primary aim of these learners in this course, the learners were required to rely on their own background knowledge and linguistic resources to craft these essays in typed format. On the topic prompt sheets, space was left for the learners to take notes if they wished. Using dictionary was not allowed throughout this independent writing practice.

Every learner completed ten typed essays in word format over the study period. Thus, the main corpora of the study consisted of 120 essays. The essay topic prompts were all taken from TOEFL iBT practice test books similar to the essay writing practices they experienced in the class. To minimize the influence of specific topics, we aimed to balance those that shared a similar academic writing register, all created under fairly similar circumstances, including themes like education, university life, school life, technology, and social activities. To enhance the internal validity, two methodological approaches were employed: topic stratification, to ensure balanced distribution of topics across participants; and sensitivity analyses, to assess robustness by systematically excluding topics exhibiting atypical linguistic patterns. These measures supported the consistency and reliability of our findings.

The participants were instructed to write their essays while observing the time limit and word length in TOEFL iBT test format. As they were getting prepared to sit the official TOEFL iBT test, such simulated practice made sense for the participants and was honoured based on the anecdotal evidence. The word length of the compositions ranged almost between 300-500 words (see Table 2). The instructor provided holistic and analytic written corrective feedback at her discretion on diverse aspects of the finished essays. She primarily focused on delivering corrective feedback related to revising, clarifying, and emphasizing various aspects of the participants' syntactic errors throughout the course. The instructor employed a structured feedback protocol to priorities specific error categories, including clause-boundary errors (e.g., comma splices, fused sentences), phrasal complexity issues (e.g., dangling modifiers, noun phrase agreement), and academic register violations (e.g., inappropriate contractions). The protocol also included a correction template featuring two strategies: (1) direct correction (e.g., "missing definite article → the") and (2) indirect correction with a metalinguistic cue (e.g., "consider reducing noun phrase density here [NP → V]").

Consequently, the corpus consisted of a total of 43,478 running words. The essays were chronologically ordered and saved in text files. Since the writing task was computerized, the students had time to proofread their writings and to correct mechanical syntactic and spelling errors so that the errors were very few in the writing samples. At the end, the topic prompts were deleted and the main texts were imported

into the syntactic and lexical analyzers. The output results of analyzers were imported to Excel and SPSS for further statistical analysis.

After analysing the texts via syntactic and lexical analyzers, the researchers subjected the results to descriptive statistics, utilizing frequency, standard deviations, and mean values on each specific component. To measure the performance of the participants over multiple time points, a repeated-measures ANOVA was conducted using SPSS *S*tatistics (version 21). This statistical methodology enabled a detailed examination of the results relevant to RQ1. A repeated-measures ANOVA provided an effective preliminary analysis given the complexity of the research question. In this study, this statistical technique was chosen to examine individual differences and variables among consistent learners over a specific defined time. Essentially, this involved calculating the mean scores differences of the learners' syntactic and lexical developments over time. This statistical technique minimizes variance and offers a more comprehensive and precise analysis. Meanwhile, to examine the relationships among variables in RQ2 and 3, Pearson product-moment correlation test was conducted.

Table 2. Number and mean of words collected per participant in different times

Name	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	MWL per essay	Total essay words
S-I	273	254	333	324	169	316	325	300	303	274	287	2,871
F-A	541	504	383	363	364	433	409	392	334	370	409	4,093
E-M	288	345	224	380	496	308	233	272	255	237	303	3,038
M-S	361	370	276	253	375	267	382	345	638	359	362	3,626
Y-A	336	534	375	352	382	409	531	387	311	436	405	4,053
H-B	282	258	258	352	182	312	288	305	289	308	283	2,834
F-D	550	502	358	289	402	299	308	356	298	401	376	3,763
F-E	305	258	405	350	389	357	302	308	358	299	333	3,331
T-H	358	405	405	455	488	488	477	425	525	368	439	4,394
G-A	358	589	281	503	399	406	560	350	351	505	430	4,302
N-F	504	257	305	308	188	190	359	421	502	251	328	3,285
R-A	504	504	504	380	452	351	358	289	257	289	388	3,888
Mean	388	398	342	359	357	344	377	345	368	341	361	43,478

Notes: T: Time MWL: Mean Word Length

Syntactic and lexical complexity measures

In order to examine and analyze syntactic complexity, L2 syntactic complexity measure (L2SCA) designed by Lu (2010) was used. L2SCA is an automatic analysis of syntactic complexity with a high reliability (.83 and 1.000 for CPC and MLS respectively), which measures writing complexity using 14 different indices explored by second language development studies (Lu, 2010). According to Lu (2010), the system is designed for analysing college-level writing data or intermediate to advanced second language writing corpus and is reliable for few-error writing texts. As described by Lu (2010), this measure is categorized into five sub-components: the length of production, sentence complexity, subordination, coordination, and particular structures. These indices, their definitions, and abbreviations are summarized in Table 3.

Table 3. Measures of syntactic complexity (Adopted from Lu, 2010)

sub-components	construct	abbreviation	definition
Production units	Clause length	LC	# words / # clauses
	Sentence length	LS	# words / # sentences
	T-unit length	LT	# words / # T-units
Sentence complexity	Sentence complexity ratio	C/S	# clauses / # sentences
Subordination	T-unit complexity ratio	C/T	# clauses / # T-units
	Complex T-unit ratio	CT/T	# complex T-units / # T-units
	Dependent clause ratio	DC/C	# dependent clauses / # clauses
	Dependent clauses per T-unit	DC/T	# dependent clauses / # T-units
Coordination	Coordinate phrases per clause	CP/C	# coordinate phrases/# clauses
	Coordinate phrases per T-unit	CP/T	# coordinate phrases/# T-units
	Sentence coordination ratio	T/S	# T-units/# sentences
Particular structures	Complex nominals per clause	CN/C	# complex nominals/# clauses
	Complex nominals per T-unit	CN/T	# complex nominals / # T-units
	Verb phrases per T-unit	VP/T	# verb phrases / # T-units

This L2 syntactic complexity analyzer identifies the occurrences of nine units of syntactic structures. Table 4 demonstrates syntactic definitions and some related instances from the corpus of this study.

Table 4. Instances of syntactic structures from essays of the present study

INDICES	Description	Examples
Word	A sequence of letters that are bounded by white space.	*in, addition, trusting, and, giving, appropriate, degrees, of, freedom, to, children,…*
Sentence	A sentence is a group of words delimited with punctuation marks (.,?, !, ”,).	*What makes parents special from other human beings is the unconditional love and the great attention they devote to their kids. (21 words)*
Clause	A syntactic structure with a subject and a finite verb.	*People suffer from traffic jam.*
Dependent clause	A finite clause that is a nominal, adverbial, or adjective clause.	It happens *when you want to get rid of a bad feeling. While old houses have not strong infrastructures* new apartments are well-designed.
T-unit	An independent clause and any clauses dependent on it.	*The same is about sporting in silence, somehow it may seem tiring for me.* (2 T-unit).
Complex T-unit	A T-unit that includes a dependent clause.	Nowadays people do not have much time (T-unit) *because they are busy with their routine life* (independent clause).
Coordinate phrase	Adjective, adverb, noun and verb phrases connected by a coordinating conjunction.	Some may complain about the lack of greenery in stylish apartments, *yet* it is an advantage for some busy people.
Complex nominals	(i) Nouns with modifiers, (ii) Nominal clauses, (iii) Gerunds and infinitives that function as subjects.	(i) *A traditional house, sealed windows* (ii) I believe that *they keep balance in their give and take process* (iii) *Choosing the right person, Providing someone requests*
Verb phrase	A finite or non-finite verb phrase that is dominated by a clause marker.	*Make them more curious* *To increase their intelligence*

To evaluate the learners' syntactic and lexical relationships, nine lexical indices were selected which were considered the best indicators based on a review of literature (see e.g., Bulte´ and Housen, 2012; Lu, 2012). To better illuminate our findings, we employed three lexical analyzers. Accordingly, Lexical Complexity Analyzer indices (Ai & Lu, 2010; Lu, 2012), Coh-Metrix (Graesser et al., 2004; McNamara & Graesser, 2012), and VocabProfile (Cobb, 2000) were selected to account for lexical density, diversity, and sophistication. These measures, definitions and software specifications are provided in Table 5. All instruments were valid and reliable tools. For instance, Coh-Metrix has demonstrated high reliability, with a reported coefficient of 0.92 in texts of specific genres (McNamara & Graesser, 2012). VocabProfile is also a reliable measure of lexical complexity with reliability of more than 0.75 for its different indices (Abbasian & Shiri Parizad, 2011). Lexical Complexity Analyzer also correlates strongly with the raters' judgments of the quality of ESL learners' oral narratives from

moderate to high (r = 0.53 to 0.76) for lexical density, diversity, and sophistication (Lu, 2012).

Table 5. Indices and software specifications to measure lexical complexity

Lexical Complexity	Indices	Definition	Software applications
Density	Lexical Density (LD-LCA)	Content word ratio	Lexical Complexity Analyzer
	Lexical Density (LD-VP)	Content word ratio	VocabProfile
Diversity	Uber Index (Uber)	The proportion of the squared number of log to the whole number of log in the text.	Lexical Complexity Analyzer
	Squared Verb Variation (SVV)	The proportion of the squared number of verb types to the whole number of verbs in the text.	Lexical Complexity Analyzer
	Measure of Textual Lexical Diversity (MTLD)	The average length of sequential word strings in a text which maintain a given TTR value.	Coh-Metrix 3
	Vocabulary Diversity (Vocd-D)	A mathematical transformation of the standard type-token ratio (TTR) which reduces the intervening impacts of text length and indicates the degree of words' repetition in a text.	Coh-Metrix 3
Sophistication	Academic Word Length (AWL)	A list of 570 frequent words in an academic context.	VocabProfile
	Beyond-2000(B-2000)	The Beyond-2000 values calculated by subtracting K1 and K2 ratios from 100%.	VocabProfile
	Content Word Log Frequency (LogF)	The average of the log frequency of content words in the text	Coh-Metrix 3

Results

To examine the process of syntactic development, we subjected the results to descriptive statistics, utilizing the mean and standard deviation for each specific component. In order to determine whether significant development occurred over time with regard to the syntactic indices, a repeated-measures ANOVA was performed. Before conducting repeated measures ANOVA, normality of the data was assessed. Almost all indices met the assumption of approximate normality. Table 6 presents the descriptive statistics for the 14 syntactic indices (see Tables 3 and 5 for abbreviations).

Table 6. Means (M) and Standard Deviations (SD) of syntactic complexity at time intervals

IND/T		T2	T3	T4	T5	T6	T7	T8	T9	T10
MLC	M	9.48	9.67	10.48	10.23	11.47	10.28	10.55	12.24	11.25
	SD	.88	.50	1.24	2.30	1.96	1.64	2.25	1.96	1.47
MLS	M	17.29	17.56	18.13	22.62	20.30	21.20	22.07	22.25	25.22
	SD	3.12	3.53	2.22	1.92	2.19	4.01	4.38	4.12	3.78
MLT	M	14.71	16.21	16.32	20.37	17.99	19.60	19.37	19.69	22.87
	SD	2.27	2.29	1.01	2.95	2.46	3.18	1.99	3.89	3.37
C/S	M	1.89	1.79	1.84	2.09	1.86	2.14	2.17	1.67	2.30
	SD	.24	.43	.31	.35	.27	.48	.67	.27	.35
CT	M	1.55	1.63	1.57	1.88	1.67	1.70	1.89	1.60	2.16
	SD	.17	.30	.27	.18	.39	.23	.38	.23	.09
CT/T	M	.43	.44	.46	.51	.47	.56	.55	.43	.69
	SD	.08	.17	.18	.12	.16	.13	.15	.16	.11
DC/C	M	.33	.34	.37	.41	.37	.43	.42	.32	.50
	SD	.07	.09	.10	.07	.11	.03	.10	.08	.02
DC/T	M	.51	.58	.61	.74	.65	.83	.83	.53	1.08
	SD	.16	.25	.26	.20	.36	.16	.31	.19	.07
CP/C	M	.25	.32	.31	.34	.33	.28	.30	.37	.26
	SD	.07	.15	.17	.21	.11	.04	.20	.24	.14
CP/T	M	.35	.49	.51	.53	.51	.49	.47	.48	.42
	SD	.11	.21	.19	.37	.17	.12	.25	.27	.33
T/S	M	1.12	1.09	1.12	1.17	1.14	1.12	1.14	1.05	1.11
	SD	.06	.09	.10	.10	.13	.12	.15	.07	.13
CN/C	M	1.12	1.10	1.12	1.48	1.11	1.26	1.54	1.71	1.19
	SD	.16	.25	.29	.45	.36	.42	.37	.52	.34
CN/T	M	1.43	1.77	1.80	2.62	1.76	2.36	2.63	2.67	2.55
	SD	.26	.35	.28	.72	.41	.76	.43	.69	.71
VP/T	M	2.23	2.36	2.41	2.71	2.69	2.85	2.79	2.43	3.19
	SD	.46	.52	.20	.19	.54	.40	.45	.40	.44

The next issue was to consider Mauchly's Test for Sphericity. Output of Table 7 shows Mauchly's test for syntactic data. Based on the important column containing the significance values which are less than .05, we accept the hypothesis that the variances of syntactic indices differences between times are significantly different. In other words, the assumption of sphericity has been violated, and Epsilon values are less than .75 (Field, 2018). Then, we used Greenhouse-Geisser and Huynh-Feldt corrections. As with the Greenhouse-Geisser correction, the Huynd-Feldt correction estimates epsilon in order to correct the degrees of freedom of the *F*-distribution.

Table 7. Results of Mauchly's Tests of Sphericity

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon	
					Greenhouse-Geisser	Huynh-Feldt
MLC	.000	67.388	11	.031	.409	.641
MLS	.000	11.245	11	.021	.533	.666
MLT	.000	37.154	11	.000	.407	.625
CS	.000	21.245	11	.000	.350	.564
CT	.001	36.254	11	.000	.351	.662
CTT	.000	37.225	11	.002	.409	.111
DCC	.000	57.524	11	.000	.409	.625
DCT	.001	63.945	11	.000	.325	.526
CPC	.000	67.254	11	.000	.721	.564
CPT	.000	51.245	11	.031	.666	.652
TS	.000	13.645	11	.021	.533	.124
CNC	.003	14.245	11	.031	.401	.625
CNT	.000	18.212	11	.001	.214	.256
VPT	.000	65.245	11	.000	.414	.635

Table 8 summarizes the results of repeated measures ANOVA, using Greenhouse-Geisser measure. In a repeated-measures ANOVA, the effect size can be calculated via the partial eta squared (η^2_p). In this case, the F result indicated a significant difference between time and mean scores of MLS, MLT, DC/T, CN/T, and VP/T (see Table 4 for abbreviations) development indices: (MLS, $p < .001$, $\eta^2_p = .521$; MLT, $p<.001$, $\eta^2_p = .565$; DC/T, $p< .001$, $\eta^2_p = .427$; CN/T, $p < .001$, $\eta^2_p = .547$; VP/T, $p < .001$, $\eta^2_p = .467$). The results also revealed positive (but non-significant) differences between time and C/T ($p = .074$, $\eta^2_p = .427$) and DC/C ($p = .084$, $\eta^2_p = .400$) scores. No significant difference was observed between time and mean scores of the seven indices (Table 8).

Table 8. Repeated-measures ANOVA results for syntactic complexity

Indices	Type III Sum of Squares	Mean Square	F	Sig.	Partial Eta Squared
MLC	35.503	13.800	1.728	.224	.302
MLS	365.508	140.719	4.357	.035	.521
MLT	283.316	94.748	5.187	.016	.565
CS	2.597	1.148	2.513	.132	.386
CT	1.741	.584	2.986	.074	.427
CTT	.305	.100	1.929	.178	.325
DCC	.151	.044	2.669	.084	.400
DCT	1.472	.483	3.301	.046	.452
CPC	.077	.030	.464	.688	.104
CPT	.266	.102	.869	.474	.178
TS	.055	.020	.746	.536	.157
CNC	1.818	.748	2.374	.139	.372
CNT	7.969	2.726	4.828	.021	.547
VPT	4.633	1.366	3.502	.041	.467

To provide a richer and deeper understanding of developmental pathways-a core tenet of DST-the data for the significantly developed indices were imported into Excel. Figure 1 illustrates the developmental trajectories of the five syntactic indices over time. As shown in the graph, the learners demonstrated overall growth in the MLT, MLS, DC/T, CN/T, and VP/T across the data collection period. To prevent visual clutter and overlapping lines, only the developmental patterns of six learners were included in the graph. The graph indicates that the learners' syntactic development followed incremental yet fluctuating patterns over time. Notably, the average VP/T scores exhibited less fluctuation compared to the other indices. Additionally, individual differences were evident, with each learner displaying a unique developmental trajectory.

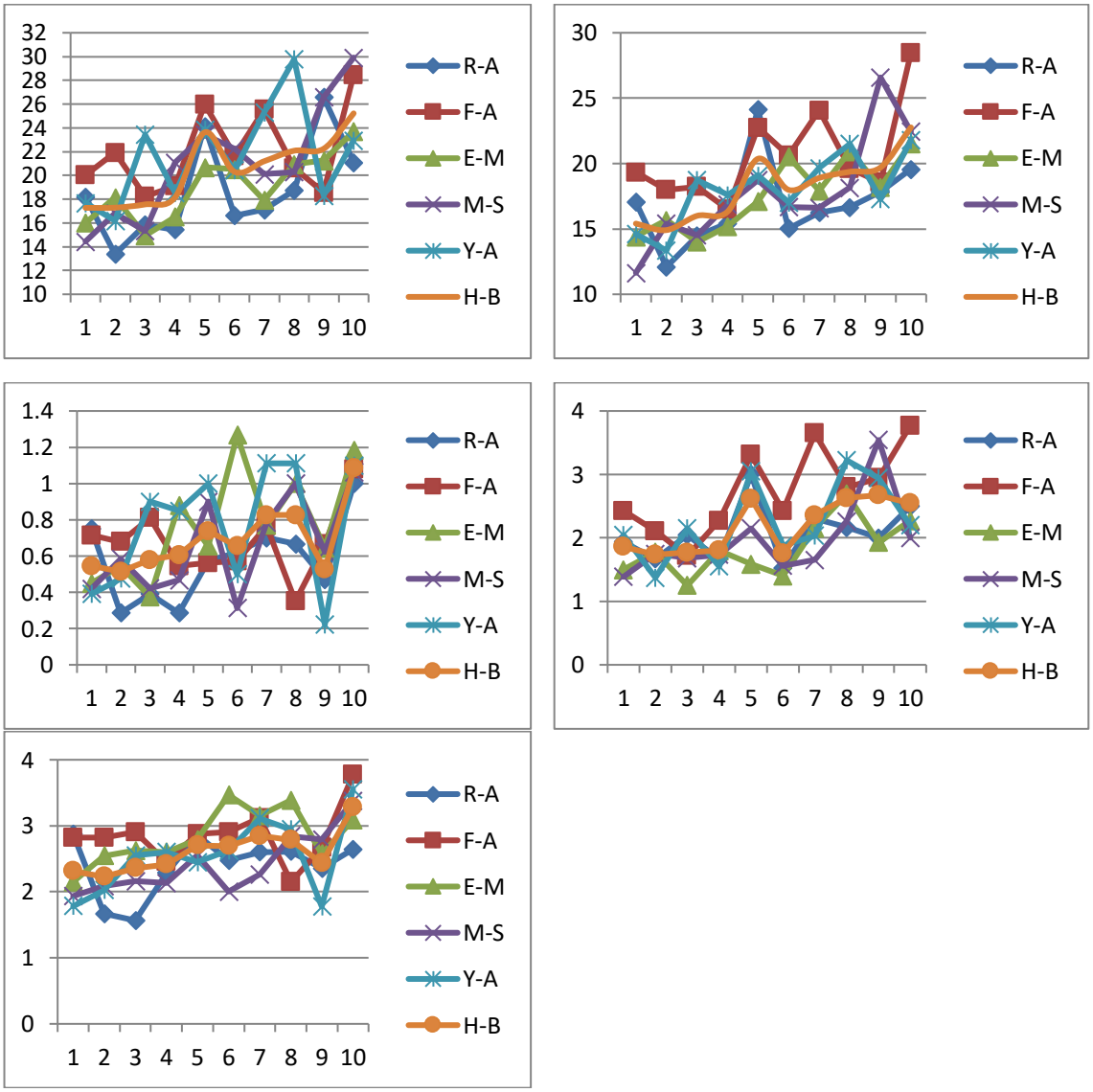

Figure 1. Learners' syntactic development on MLT, MLS, DC/T, CN/T, and VP/T. (The graphs are arranged from left to right, and up to down, respectively.)

To examine the relationships among the dimensions of syntactic complexity, correlation analyses were conducted for each pair of indices. Table 9 presents the correlation coefficients among these measures.

Table 9. Correlations between syntactic complexity measures

	MLC	MLS	MLT	C/S	C/T	CT/T	DC/T	CP/C	CP/T	T/S	CN/C	CN/T	VP/T
MLC													
MLS	.492 .200												
MLT	.828* .042	.761 .067											
C/S	-.477 .208	.527 .181	-.011 .493										
C/T	-.391 .258	.427 .236	.188 .381	.854* .033									
CT/T	-.782 .059	-.018 .488	-.302 .311	.773 .063	.858* .031								
DC/T	-.421 .240	.253 .341	.136 .414	.720 .085	.966** .004	.852* .033							
CP/C	.839* .038	.602 .142	.591 .147	-.246 .345	-.458 .219	-.793 .055	-.582 .152						
CP/T	.830* .041	.467 .214	.669 .108	-.334 .292	-.289 .318	-.714 .088	-.271 .329	.783 .059					
T/S	-.452 .223	.499 .196	-.156 .401	.915* .015	.576 .155	.572 .157	.383 .262	-.050 .468	-.353 .280				
CN/C	.942** .008	.553 .167	.774 .062	-.362 .275	-.337 .290	-.772 .063	-.361 .275	.882* .024	.963** .004	-.340 .288			
CN/T	.912* .016	.703 .093	.891* .021	-.170 .392	-.094 .440	-.591 .147	-.133 .416	.813* .047	.926* .012	-.220 .361	.967** .004		
VP/T	.313 .304	.176 .388	.634 .125	-.055 .465	.449 .224	.188 .381	.566 .160	-.200 .374	.203 .372	-.423 .239	.208 .368	.363 .274	

*. Correlation is significant at the 0.05 level (2-tailed). **. Correlation is significant at the 0.01 level (2-tailed).

To interpret the results of syntactic correlation, correlations among the indices and between the subcomponents were considered. MLC was most strongly correlated with MLT (r = .828), CP/C (r = .839), CPT (r = .830), CN/C (r = .942), and CN/T (r = .912). MLT was correlated with CN/T (r = .891); CS was correlated with C/T (r = .854) and T/S (r = .915); C/T was correlated with CT/T (r = .858), and DC/T (r = .966); C/TT was correlated with DC/T(r = .852); CP/C was correlated with CN/C (r = .882) and CN/T (r = 813); CP/T was correlated with CN/C (r=.963), and CN/T was strongly correlated with CP/C (r = .813), CP/T (r = .927), and CN/C (r = .967). There were also negative (but not significant) correlations between some indices such as MLC and CT/T or MLT and CT/T. Interestingly, MLS showed no correlation with other indices.

An interesting pattern in this correlation matrix was high correlations among the measures of the same categories, as we expected. For example, in the first category, MLC and MLT indices were highly correlated, but these indices showed low correlation with MLS. C/T, CT/T, and DC/T indices were highly correlated. In coordinate category, there was not any correlation between CP/C, CP/T, and T/S indices. CN/T and CN/C were strongly correlated, but these exhibited low correlation with VP/T.

Furthermore, correlation coefficients of syntactic subcomponents indicated that particular structures had a positive association with the subcomponent groups of coordination and length of units. However, the subordinate group revealed a negative (but non-significant) relationship with both coordination and particular structures groups. Similarly, lengths of units exhibited a negative (but non-significant) relationship with both subordination and coordination groups.

To estimate the relationships between the dimensions of syntactic and lexical constructs, the correlations between some selected indices from both syntactic and lexical measures were calculated (Table 10).

Table 10. Correlations among syntactic and lexical components

Syntactic Indices	Lexical density		Lexical diversity				Lexical sophistication		
	LD-LCA	LD-VP	MTLD	Uber	SVV	Voc-D	AWL	B-2000	LogF
MLC	.529	.425	.845*	.793*	.065	.190	.495	-.315	-.246
Sig.	.360	.514	.036	.039	.140	.380	.198	.303	.345
MLS	-.287	-.254	.059	.027	.821*	.854*	-.067	-.396	.612
Sig.	.640	.562	.463	.483	.044	.045	.470	.255	.138
MLT	.014	.214	.630	.643	.869*	.354	-.007	-.55	.301
Sig.	.983	.862	.121	.121	.028	.424	.496	.484	.312

*. Correlation is significant at the 0.05 level (2-tailed).

As Table 10 illustrates, MLC, MLS, and MLT revealed highly positive correlation with only lexical diversity indices. Among these, MLC was most strongly correlated with MTLD (r=.845) and Uber (r=.793). MLS was correlated with SVV (r=.821) and Voc-D (r=.854). MLT showed a significant correlation with SVV (r=.869). Syntactic indices indicated no significant correlation with lexical density and sophistication.

To study the data descriptively, some samples from the learners were analyzed. To add more details, one from the least developed and one from the most developed texts were selected for more descriptions. The first two extracts are S-I's written text in time intervals 2 and 5. The second two extracts are from F-A's written text in time intervals

2 and 10. Based on the analysis of the raw data from the 12 learners, S-I showed the least improvement in overall syntactic development, whereas F-A demonstrated the greatest improvement. Table 11 presents the syntactic development of these two learners over time, focusing on five indices: MLT, MLS, DC/T, CN/T, and VP/T. Quantitative analysis of the syntactic development of the 12 learners, measured using indices such as clause complexity and T-units, revealed considerable variation in their rates of improvement. Participant S-I demonstrated minimal growth, whereas participant F-A exhibited the most substantial gains compared to the other learners. For example, F-A's mean length of T-unit (MLT) increased by 52% between the first- and final-time intervals, while S-I's MLT increased by only 12%. In addition, F-A demonstrated the use of more complex syntactic structures, including passive constructions, whereas S-I's writing predominantly relied on simple subject–verb–object (SVO) patterns (see the extracts).

Table 11. Learners' syntactic development over time

Name	INDIC	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
F-A	MLS	20.03	21.91	18.23	19.10	26.11	21.65	25.56	24.63	20.55	28.46
S-I		13.36	15.85	15.42	16.14	16.63	17.10	18.75	17.82	19.07	18.20
F-A	MLT	19.32	18.00	18.23	16.50	22.75	20.61	24.05	19.60	21.66	29.46
S-I		17.06	12.09	14.47	15.42	21.14	15.04	16.25	16.66	17.82	19.57
F-A	VP/T	2.82	2.82	2.90	2.45	2.87	2.90	3.11	2.15	2.66	3.76
S-I		2.87	1.66	1.56	2.28	2.85	2.47	2.60	2.61	2.35	2.64
F-A	DC/T	0.71	0.67	0.80	0.54	0.56	0.57	0.76	0.35	0.66	1.076
S-I		0.35	0.28	0.39	0.28	0.57	0.61	0.70	0.66	0.47	0.71
F-A	CN/T	2.42	2.10	1.71	2.27	3.31	2.42	3.64	2.80	2.94	3.76
S-I		1.93	1.66	2.04	1.66	2.70	1.52	2.30	2.16	2.00	2.50

Excerpt 1:

> In the last century, beside the growth of technology, the medical case has also developed. We have been witnessing many discoveries in medicine. It has helped humans to cease the spread of epidemic disease. Nowadays, people live longer and life expectancy has increase. In my point of view, the disadvantages of improving medical care negligible in comparison with its advantages.

Excerpt 2:

> However, in my opinion, it is important for children to spend their free time doing other activities because activities such as being with friends in society and playing in a group can help children to be more social, and also increase their intelligence and make them more curious.

In Extract 1, S-I produced a relatively short written text, with an MLT of 10, an MLC of 10, and an MLS of 12. Nevertheless, she demonstrated some complexity at the phrasal level, as evidenced by the use of a complex nominal phrase ("the spread of epidemic disease"). In Extract 2, S-I's writing showed an increase in syntactic complexity overall. She used longer T-units, sentences, and clauses, indicating a shift toward more complex sentence structures.

Nonetheless, these syntactic features do not necessarily indicate greater complexity. As Dynamic Systems Theory (DST) emphasizes, language development emerges from nonlinear interactions between multiple subsystems (Verspoor et al., 2011; de Bot et al., 2007). It's essential to take all indicators into account. Consistent with Verspoor et al. (2011), we analyzed complexity as a dynamic equilibrium between syntactic and lexical subsystems rather than isolated features.

Excerpt 3:

> Imagine celebrations like birthdays and weddings without music, lacking the sense of joy and happiness. The same is about sporting in silence; somehow it may seem tiring for me. Thereby, it seems most of activities we used to enjoy would not sense the same unless music be present in the background.

Excerpt 4:

> Having considered the experience, the results and the major as well, every teacher may have chosen a specific teaching method for his classes. However, the student's capabilities in different classes ought to be observed. While the supporters of student-centred education system are growing, I personally prefer teacher-centred type, which I find more comfortable and beneficial and efficient.

Extracts 3 and 4 reveal F-A's writing development in two different time intervals. In Extract 3, MLT = 12.75, MLC = 8.5, and MLS = 17. In Excerpt 4, MLT = 19, MLC = 11.6, and MLS = 19, which indicates that F-A used longer T-units in her writing at the end of the program. Careful examination of sentence complexity also reveals how F-A rightly joined two dependent clauses and one independent clause to make a compound-complex sentence at the end of the program. She also used a perfect form of gerund and more collection words (e.g., *the student's capabilities in different classes*) to construct a complex noun phrase.

Discussion

How does the syntactic complexity of EFL learners' L2 writing develop over time?

Considering writing competence as a multi-component construct (Norris & Ortega, 2009), the present longitudinal study revealed complex findings in syntactic development. Among 14 indices of syntactic variables, five indices showed significant changes, including two measures of length of production: MLS and MLT; one coordinate measure: DC/T; and two particular structures measures: CN/T and VP/T. No significant development was observed in the remaining nine indices. These results were generally consistent with previous studies (e.g., Kyle, 2016; Lu, 2010 & 2011; Salsbury, 2000). For instance, Kyle (2016), through an analysis of Salsbury's (2000) written data, identified significant development in four indices of syntactic complexity. However, the findings of this study were inconsistent with several previous studies that reported significant development in other components of syntactic complexity. For example, in the current study, the C/S index did not show significant development over time. This result contrasts with Lu's (2011) finding, which reported a negative relationship between C/S and school year. In the subordination group, only one index; DC/T, which measures the amount of clausal subordination in a text, revealed significant difference over time. This finding supports Homburg (1984) who showed a significant positive relationship between DC/T and proficiency, but it is inconsistent with Lu (2011) who demonstrated a negative relationship between DC/T and proficiency. Meanwhile, CT/T did not have a positive difference in development. According to Kyle (2016), a positive relationship between proficiency and CT/T suggests that learners use more independent/dependent clause combinations, but would not be able to determine the number or type of dependent clauses. The other two indices, C/T and DC/C, showed positive developmental trends, although these changes were not statistically significant—similar to the findings reported by Casanave (1994). However, this result aligns with those of Lu (2011), who also did not find significant developmental differences in C/T over time. Lu (2011) also found a negative relationship between DC/C and school level, suggesting that writers use fewer dependent clauses as their language proficiency increases. None of coordination indices in this study revealed a positive significant difference over time. In contrast,

Lu (2011) found a positive relationship between CP/C and language development, although this relationship was only significant after four years of studying English. Of the five studies reviewed by Wolfe-Quintero et al. (1998), only one reported a significant difference between language proficiency and T/S index as coordination structure. From three indices in particular structure, two indices, namely CN/T and VP/T, showed significant difference over time. Unlike Lu (2011), this study did not find a significant difference in CN/C over time. In Lu's (2011) study, CN/T discriminated between the first year and years 2-4, while CN/C discriminated between all levels except years 3-4. The study also showed a significant development in VP/T index which was not consistent with Lu (2011) who found no relationship between VP/T index and proficiency. This suggests that more proficient learners use more verb phrases as their writing develops over time. This result is also consistent with Biber et al.'s (2011) findings, regarding the importance of phrasal complexity over clausal complexity in L2 writing proficiency.

Descriptive analysis of the learners' essays also revealed that the learners' lexical and syntactic complexities were improved. The reason for such development was evident in their elaboration ability. For example, the analysis of the learners' writing development revealed that the frequency rates of particularly simple and vague nouns and public verbs were significantly higher at the commencement than at the end of the course. For example, the rates of the occurrence of spoken and informal lexical features, vague nouns (e.g., people, society, things,) and public verbs (e.g., spend, pay, affect, become, think…) were higher in essays which written at the beginning of the course. The findings of this study confirmed that individuals exhibited a significant degree of variability and non-linearity in their writing evolvement. The possible explanation for the learners' variations and non-equilibrium in syntactic developmental patterns is the existence of co-adaptive interactions between the sub-systems of dynamic systems which may lead to the spontaneous formation of trade-off relationship. According to Larsen-Freeman (2019), DST characterizes system as spatially–temporally situated which can be achieved through iteration and co-adaptation. If we regard the results from this perspective, undeveloped syntactic indices presumably mirror attractor states as a result of co-adaption. Developing or regressing in one section of a system is in response to other section of the system.

What are the relationships among the sub-components of syntactic complexity (i.e., length of units, sentence complexity, subordination, coordination, particular structure)?

The findings of this section demonstrated that syntactic indices and subcomponents were correlated. This is in line with a number of studies demonstrating positive relationships between MLS or MLT and language proficiency (Lu, 2011; Ortega, 2003; Wolfe-Quintero et al., 1998). According to Lu (2011), these correlations contribute to determination of measures which should be considered in describing learner proficiency. Additionally, certain measures revealed similar patterns of development. For example, MLC was strongly correlated with CP/C. The relations between the subcomponents support DST principle in which groups of syntactic measures interact.

We may explain such positive correlation among syntactic indices by the fact that syntactic indices are not separate constructs, but multifaceted and interconnected and thus not reducible to an individual variable. On the basis of DST, it could be possible to estimate how syntactic complexity is relevant in complexity in L2 writing. Thus, it is plausible to look at writing complexity not as a single variable but as a complex one. This result reflects multidimensional nature of L2 complexity in which a set of interrelated components such as sentence, phrase, and clause constitute the complexity of syntax (Bulte´ and Housen, 2012).

Considering non-linearity and individuality, it could be inferred from our analysis that on average, the learners significantly increased several syntactic indices over time. However, the average of VP/T exhibited less fluctuation than DC/T and CN/T which revealed relatively more fluctuation than others. However, the mean scores of MLS and MLT indicated relatively linear fashion at the beginning and at the end of the program. Individual differences also demonstrated varied patterns of difference over time, and each learner had her own path of development. The participants' TOEFL iBT writing scores revealed that two of them got low score compared to their classmates. However, at the end of the course they developed their lexical and syntactic complexities similar to the more proficient ones. Nevertheless, they showed relatively more fluctuation than the others. These findings support the claims made by Bulte´ and Housen (2018) regarding a high degree of variability by individuals deviating from

the mean group trends. The plausible justification in the case of less or more fluctuation concerns the stability of the system's state.

What is the relationship between EFL learners' syntactic and lexical development?

Our discussion concerns the inquiry to what extent syntactic and lexical complexity are in(ter)dependent. Unlike non-supportive development of different aspects of lexical development in terms of density, diversity and sophistication among high proficiency learners (Kalantari & Gholami, 2017), syntactic development confirmed a parallel relationship with lexical development but only with lexical diversity. Mean length of T-units, mean length of sentence and mean length of clause indicated a positive correlation with lexical diversity. In other words, when the participants used more complex sentences, they were more likely to use different words, and these words in turn contributed to production of more complex sentences. Consequently, vocabulary variation would be a precursor for sentence complexity and vice versa. However, syntactic production length showed no significant correlation with lexical density and sophistication. This finding is inconsistent with Verspoor and van Dijk (2011), who reported a positive and highly significant relationship between the development of lexical sophistication and syntactic complexity.

The following interpretations could be made to account for interdependency of syntactic and lexical diversity development as was the case in this study. There are two distinct viewpoints about this issue: one is proposed by researchers such as VanPatten (1996) and Vercellotti (2015) who confirm the view that L2 learners' attentional pools and processing capacity are constrained by limited recourses, making it difficult for learners to attend simultaneously to disparate aspects of complexity. According to this view, developments at one side of complexity may occur at the expense of developments on the other side of complexity. However, a different approach proposes the view that learners can simultaneously access multiple non-competitional and attentional mechanisms by increasing the cognitive demands of tasks (Robinson, 2003). According to Robinson's (2003) findings, under appropriate condition, development in one dimension may facilitate the emergence of other aspects of language output. Based on this issue, the possible justification for the syntactic and lexical diversity correlation may be due to the proficiency level of the learners; their

lexical and syntactic skills reinforced each other. On the other hands, employing intricate syntactic structures and sophisticated vocabulary enhances cognitive demands, necessitating that learners strike a balance to maintain fluency and accuracy in their language use. This balancing act may account for the complexities that have been observed. In summary, while cognitive load theory focuses on optimizing how information is presented to enhance learning, DST emphasizes the evolving nature of language acquisition. Based on the DST framework, the presumed reason for this is that unlike beginners, advanced learners may have dual concentration on two different aspects of language simultaneously when using their cognitive resources. These advanced learners could craft a joint operation for syntactic and lexical diversity at the same time and the two subsystems continued to co-adapt to each other in a parallel form. Thus, syntactic complexity and lexical richness could have cross-fertilized one another.

However, the reason for non-supportive development of syntactic and other aspects of lexical development could be due to the context of our study. We contend that, among various indices of vocabulary knowledge, lexical sophistication is particularly susceptible to improvement over a short period and has therefore received greater attention in instructional contexts. It is likely that the students in this study focused on acquiring a limited set of sophisticated words, such as those commonly found in GRE word lists. These learners may have intentionally inserted such words into their essays in an effort to enhance their lexical sophistication, which may not necessarily reflect broader or parallel development in their overall linguistic repertoire.

Conclusion and implication

This research yielded several substantive findings with significant implications for L2 writing development and assessment research. First, the syntactic indices including MLS, MLT, DC/T, VP/T, and CN/T were shown to develop significantly over time. This provides strong evidence indicating that as language learners become more proficient, they tend to write more T-units, longer sentences, dependent clauses per T-unit, more complex nominal, and verb phrases over time. Second, positive correlations were found among syntactic indices and subcomponents, revealing that certain syntactic indices are not separate constructs, but are interconnected. Finally, a significant

correlation was identified between lexical diversity and sentence complexity, indicating a parallel developmental relationship between lexical and syntactic growth.

The current study presents various theoretical and pedagogical connections to the field of studies in language development and assessment. Theoretically, it clarifies the intricate inherent nature of L2 writing from DST perspective. That is, it substantiates the issue that second language writing, is a complex construct that involves different underlying components and structures. In order to conduct a thorough assessment, it is essential to first analyze the structure of writing in a comprehensive manner. As noted by Li and Zheng (2024), DST views languages and language learners as complex and dynamic systems consisting of multiple components that interconnect and interact with each other to generate some overall states at a specific moment (Larsen-Freeman & Cameron, 2008). Pedagogically, the findings of this study may have practical applications in schools, language institutes, and universities. EFL teachers can benefit from these systematic findings by gaining insights into how patterns of development and individual variation manifest over time, as well as how different constructs of writing—such as syntactic complexity and lexical diversity—are interrelated. Such knowledge may inform instructional practices aimed at supporting learners' writing development more effectively. This suggests that language learning does not follow an additive pattern; instead, it can vary greatly and show individual progressions and unpredictable evolving system. In Larsen-Freeman's (2014) words, it is an ever-developing and open system where patterns emerge from the continuous self-organization among multiple components at multiple scales.

As noted by Li and Zheng (2024), regarding the language system, this perspective rejects the reductionist view of reducing the complex situation to a universal principle that removes all the "noises" in a "grand-sweep" way (Lowie & Verspoor, 2015). Hence, process-oriented analysis of L2 writing development could assist EFL instructors to identify the individuals' variation in syntactic complexity achievement. It is also important to both look at group trend and individual developmental trajectories for a full understanding of L2 writing development. As shown in graphic lines the learners exhibited varied patterns of evolution in their writing. However, individual differences revealed distinct developmental patterns, with each learner following her own unique trajectory. Furthermore, material developers can get advantage of the results of this

study since findings of the present study may identify improvement in preparing materials for writing. As a case, an awareness of the trade-off relationship in performance can help them to make a balance between the syntactic and lexical components and sub-components in materials development.

Last but not least, proposing novel forms of L2 complexity measurement can potentially characterize how variations should be taken into account in lexical and syntactic complexity evolution. Another noteworthy methodological connection to language assessment is that employing multiple analyzers and indices for syntactic and lexical development in an extended period of time is fruitful and promising. Employing triangulated measures to analyze syntactic and lexical construct can assist researchers and teachers to evaluate them far more quickly, automatically, and reliably. These automated measures can be used as a diagnostic tool to assess students' syntactic and lexical development in order to improve their writing skill. In validating testing procedures and their interpretation, McNamara (2014) reminds us to have plenty of evidence in order to judge reliability and validity, cautioning against unsound or faulty interpretations.

As with most studies, the present study has several limitations. Despite the considerable range of indices employed and the use of reliable, statistical techniques and software, these tools may still be too limited to fully and precisely capture the complexity of L2 learners' syntactic production. For example, advanced statistical techniques such as structural equation modeling could be used to analyze the dynamic interrelationships among variables and model their developmental trajectories. Future research may benefit from incorporating such techniques to provide more robust and comprehensive insights. A similar limitation applies to the analysis of lexical complexity subcomponents, which may require more nuanced methods to capture their developmental patterns. The insights obtained from the study raise some interesting questions on how other writing elements such as cohesion and coherence interact with syntactic and lexical complexity in L2 writing development and assessment. Finally, the generalizability of the present findings is limited by the relatively homogeneous sample of participants, who were predominantly at intermediate to advanced levels of English proficiency. Future studies could benefit from including a more diverse sample, spanning a wider range of proficiency levels

(e.g., novice to expert), to provide deeper insights into developmental patterns across learner populations.

Author disclosures

The authors reported no potential conflict of interest that could influence this research.

This work derives from the first author's doctoral dissertation, completed at Urmia University, Iran. No additional external funding was received for this specific publication.

The original dissertation research was conducted in accordance with Urmia University's ethical guidelines.

The authors contributed to this research and the preparation of the dissertation as follows: Reza Kalantari: conceptualization, data collection, drafting the original manuscript, and critical revision; Javad Gholami: supervision, conceptual guidance, substantive revision, and final proofreading.

Acknowledgments

The authors would like to offer special thanks to Diane Larsen-Freeman, Professor at the University of Michigan, the USA and Professor Marjolijn Verspoor the Chair of English Language and Culture Department at the University of Groningen for their very helpful and valuable comments on the earlier draft of this paper.

ORCID iDs

Reza Kalanatri https://orcid.org/0000-0003-2694-7959

Javad Gholami https://orcid.org/0000-0002-4970-1110

References

Abbasian, G. R., & Shiri Parizad, M. (2011). Validation of lexical frequency profiles as a measure of lexical richness in written discourse. *Journal of Technology & Education, 5*(3), 1–7.

Ai, H., & Lu, X. (2010, June 8–12). *A web-based system for automatic measurement of lexical complexity* [Paper presentation]. 27th Annual Symposium of the Computer-Assisted Language Consortium (CALICO-10), Amherst, MA, United States.

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

Bulte´, 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). Amsterdam: John Benjamins.

Bulte´, B., & Housen, A. (2018). Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics, 28*(2), 123–142. https://doi.org/10.1111/ijal.12196

Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing, 3*(3), 179-201. https://doi.org/10.1016/1060-3743(94)90016-7

Caspi, T. (2010). *A dynamic perspective on second language development* [Unpublished doctoral dissertation]. University of Groningen.

Cobb, T. (2000). Web VocabProfile [accessed October, 2016 from http://www.lextutor.ca/vp]

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

de Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition, 10*(1), 7-21. https://doi.org/10.1017/S1366728906002732

Ellis, N. (2007). Dynamic systems and SLA: The wood and the trees. *Bilingualism: Language and Cognition, 10*(1), 23-25. https://doi.org/10.1017/s1366728906002744

Field, A. P. (2018). Discovering statistics repeated measures ANOVA: Issues with repeated measures designs. *Discovering Statistics*. https://www.discoveringstatistics.com

Homburg, T. J. (1984). Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL Quarterly, 18*(1), 87-107. https://doi.org/10.2307/3586337

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers, 36*(2), 193-202.

Kalantari, R., & Gholami, J. (2017). Lexical complexity development from dynamic systems theory perspective: Lexical density, diversity, and sophistication. *International Journal of Instruction, 10*(4), 1-18. https://doi.org/10.12973/iji.2017.1041a

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. *The Modern Language Journal, 102*(2), 1-17. https://doi.org/10.1111/modl.12468

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics, 18*(2), 141-165. https://doi.org/10.1093/applin/18.2.141

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. https://doi.org/10.1093/applin/aml029

Larsen-Freeman, D. (2014). Saying what we mean: Making the case for second language acquisition to become second language development. *Language Teaching, 47*(1), 1-15. https://doi.org/10.1017/S0261444814000019

Larsen-Freeman, D. (2019). On language learner agency: A complex dynamic systems theory perspective. *The Modern Language Journal, 103*(S1), 61-79. https://doi.org/10.1111/modl.12536

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford University Press.

Li, H., & Zheng, Y. (2024). Complex dynamic systems theory in second language learning and teaching: A textometric review from 2008 to 2022. *Language Teaching Research Quarterly, 39*, 68-95.

Lowie, W., & Verspoor, M. (2015). Variability and variation in second language acquisition orders: A dynamic reevaluation. *Language Learning, 65*(1), 63-88. https://doi.org/10.1111/lang.12093

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. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal, 96*(2), 190-208. https://doi.org/10.1111/j.1540-4781.2011.01232_1.x

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing, 29*, 16-27. https://doi.org/10.1016/j.jslw.2015.06.003

Mancilla, R. L., Polat, N., & Akcay, A. O. (2015). An investigation of native and nonnative English speakers' levels of written syntactic complexity in asynchronous online discussions. *Applied Linguistics, 37*(5), 1-24. https://doi.org/10.1093/applin/amv012

McGrath, I. (1997). Feeding, leading, showing, throwing: Process choices in teacher training and trainer training. In I. McGrath (Ed.), *Learning to train: Perspectives on the development of language teacher trainers* (pp. 13-22). Prentice Hall.

McNamara, T. (2014). 30 years of language assessment research: Highlights from Language Testing (1984–2013). *Language Testing*, 31(4), 439–451. https://doi.org/10.1177/0265532214544403

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 188-205). IGI Global.

Naghdipour, B. (2016). English writing instruction in Iran: Implications for second language writing curriculum and pedagogy. *Journal of Second Language Writing, 32*, 81-87. https://doi.org/10.1016/j.jslw.2016.05.001

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. (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

Panahi, A., Mohebbi, H., & Larsen-Freeman, D. (2024). Diane Larsen-Freeman's fifty-year contributions to language learning and teaching: A systematic review and Diane Larsen-Freeman's personal reflection. *Language Teaching Research Quarterly, 39*, 372-425.

Polat, B., & Kim, Y. (2014). Dynamics of complexity and accuracy: A longitudinal case study of advanced untutored development. *Applied Linguistics, 35*(2), 184-207. https://doi.org/10.1093/applin/amt013

Rimmer, W. (2006). Measuring grammatical complexity: The Gordian knot. *Language Testing, 23*(4), 497-519. https://doi.org/10.1191/0265532206lt3390a

Robinson, P. (2003). Attention and memory during SLA. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 631-678). Blackwell.

Salsbury, T. (2000). *The acquisitional grammaticalization of unreal conditionals and modality in L2 English: A longitudinal perspective* [Unpublished doctoral dissertation]. Indiana University.

Schmid, M. S., Verspoor, M., & MacWhinney, B. (2011). Coding and extracting data. In M. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 39-54). John Benjamins.

Shardlow, M., Evans, R., & Zampieri, M. (2022). Predicting lexical complexity in English texts: The Complex 2.0 dataset. *Language Resources and Evaluation, 56*(3), 1-23. https://doi.org/10.1007/s10579-022-09588-2

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in the development of accuracy and complexity: A longitudinal case study on the acquisition of Finnish. *Applied Linguistics, 31*(4), 532-553. https://doi.org/10.1093/applin/amq001

VanPatten, B. (1996). *Input processing and grammar instruction: Theory and research*. Norwood, NJ: Ablex.

Vercellotti, M. L. (2015). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics, 36*(5), 1-23. https://doi.org/10.1093/applin/amv002

Verspoor, M. H, de Bot, K., & Lowie, W. (2011). *A dynamic approach to second language development: Methods and techniques*. John Benjamins.

Verspoor, M. H, Lowie, W., & van Dijk, M. (2008). Variability in L2 development from a dynamic systems perspective. *The Modern Language Journal, 92*(2), 214-231. https://doi.org/10.1111/j.1540-4781.2008.00715.x

Verspoor, M. H., & van Dijk, M. (2011). Visualizing interaction between variables. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), *A dynamic approach to second language development: Methods and techniques* (pp. 85-98). John Benjamins.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. University of Hawai'i Press.