Unpacking Reading Text Complexity: A Dynamic Language and Content Approach

Haimei Sun[1]
Teachers College, Columbia University

ABSTRACT

Reading texts may serve as a vital source of informational content and language knowledge for second language (L2) learners, which is known as the dual relevance of L2 reading (Han & D'Angelo, 2009). Text complexity, however, influences the two, determining to a great extent whether the dual outcome is attainable. Drawing on Complex Dynamic System Theory (CDST, de Bot, 2017), the present study conceives text complexity as residing primarily in the interrelationship between language and content complexity and their respective subsystems. To apply this conceptualization to text analysis, this study analyzed language complexity and content complexity of authentic science texts sampled from high school and college textbooks on four different subjects. Results show that college texts in general exhibited greater language and content complexity than high school texts, especially in terms of sentence length and the use of complex nominals. Aside from this emerging pattern, variability characterized the magnitude of difference in complexity and the manner in which the texts differed. Overall, the findings from this study attest to the mutual relationship between language complexity and content complexity embedded within authentic texts.

Keywords: text complexity, readability, CDST

INTRODUCTION

Reading texts may serve as a vital source of informational content and language knowledge for second language (L2) learners, which is known as the dual relevance of L2 reading (Han & D'Angelo, 2009). Text complexity, however, influences the two, determining to a great extent whether the dual outcome is attainable. Reading texts that are too complex may render minimal learning of either content or language. However, those that appear too elementary may not engage learners in in-depth processing of the texts, resulting in a superficial understanding of the

[1] Haimei Sun is a doctoral student in Applied Linguistics at Teachers College, Columbia University, specializing in second language acquisition. Her research interests include task-based language teaching and learning, second language reading, eye-tracking, and automatic text processing. She is currently the academic coordinator of the Center for International Foreign Language Teacher Education (CIFLTE) at Teachers College. Correspondence should be sent to E-mail: hs2700@tc.columbia.edu

content due to their reliance on schematic knowledge (Sharwood Smith, 1986). Therefore, the assessment and employment of texts of appropriate complexity in reading instruction play a pivotal role in optimizing the dual outcome of reading comprehension and language acquisition.

To assess text complexity, over 200 readability formulas have been devised (DuBay, 2004). However, such formulas have been criticized for reducing text complexity to variations of two major linguistic features: sentence length and word difficulty (e.g., Bernhardt, 1984). To fully capture the construct, other factors that have served as proxies of text complexity have run the gamut from text structure (Meyer, 1975), cohesion (Crossley et al., 2008; Graesser et al., 2004), genre (Sheehan et al., 2010), text length (Mesmer, et al., 2012), task (Valencia et al., 2014), to reader variables such as motivation and background knowledge (CCSS Initiative, 2010; Masi, 2003). However, it remains doubtful if we have come any closer to conceptually and empirically unraveling the essence of the construct. For one, text complexity is often conflated with text difficulty (Masi, 2003; Mesmer et al., 2012), leading to the inclusion of miscellaneous factors (see CCSS, 2010, for an example). Mesmer et al. (2012) argue that while text difficulty implies the learner's actual or predicted performance, serving as a *dependent* or *criterion variable*, text complexity implicates text inherent features to be manipulated as *independent* or *predictor variables*. As such, text complexity is a cause of text or processing difficulty (Bulté & Housen, 2012; Merlini Barbaresi, 2003). In addition, a lack of theoretical underpinnings might have contributed to the various conceptualizations of text complexity, including factors such as language complexity, text structure, genre, and task that are often examined in isolation and interpreted as static, as opposed to more dynamic, traits.

In view of those issues, the present study seeks to conceive text complexity as a complex dynamic system (Larsen-Freeman, 1997; Merlini Barbaresi, 2003), one that emerges primarily from the interrelationship between language and content complexity and their respective subsystems. In order to apply this conceptualization to text analysis, this study focused on hard science texts, which tend to be conceptually complex, therefore posing challenges for both first language (L1) and L2 learners who need to acquire science knowledge and essential academic language (Arya et al., 2017). Specifically, it examined how science texts sampled from high school and college textbooks differed in language and content complexity. In what follows, the study reviews relevant literature on text complexity and presents the results of text analysis, with a particular focus on the dynamic interplay between language and content complexity.

REVIEW OF LITERATURE

Theoretical Underpinnings for Conceptualizing Text Complexity

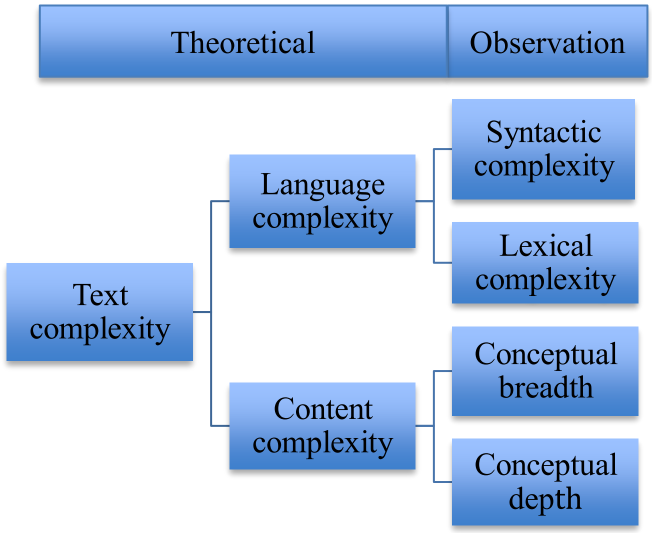
Informed by the way natural scientists investigate phenomena in their fields, applied linguists in recent years have witnessed a contemporary shift towards conceiving language, language learning and use as a complex system (de Bot, Lowie, & Verspoor, 2007; Larsen-Freeman, 1997). By merging Complexity Theory and Dynamic System Theory into Complex Dynamic System Theory (CDST) (de Bot, 2017), Larsen-Freeman (1997) contends that language is a complex adaptive system consisting of different interdependent subsystems, such as phonology, lexicon, syntax, semantics and pragmatics, and that the behavior of the system as a whole results from the aggregate of the local interactions of its subsystems. She further expounds that each time language is used, it "sets in motion a process, which may lead to change at the global level" (p. 148), and that the use of a particular linguistic form follows its function. In a similar vein, written text as one form of language use can also be construed as a complex dynamic system that encapsulates the writer's intended thoughts through the vehicle of language. Accordingly, the complexity of a text resides naturally in the intricately dynamic workings of the language deployed and the thoughts conveyed.

Indeed, the relationship between language and thought, or code and content, has long piqued the interest of first language (L1) and second language (L2) researchers (e.g., Cromer, 1988; Robinson, 2003). Notably, the Cognition Hypothesis, first proposed by Cromer (1988) to hypothesize a deterministic and interdependent relationship between conceptual and language development in child L1 acquisition, was extended to adult task-based L2 development by Robinson (2003). In particular, Robinson (2001) proposed the engineering of cognitive complexity in task design features, postulating a positive correlation between task complexity and learner language. Han and Kang (2018), however, argue that it is task-induced complex thoughts that may drive the complexity of learner language. They also aptly point out that the association between language and thought may be compromised in L2 learners due to their limited L2 proficiency; nonetheless, a stronger link between the two should bear out in native speakers' oral and written productions. In fact, Givón (1985) has long contended that "[a]ll other things being equal, structural complexity tends to accompany functional complexity in syntax" (p. 1021), suggesting that more complex thoughts entail more complex linguistic means to express them. It thus seems justifiable to reason that the language and thought/content complexities in authentic texts are likely to go hand in hand, functioning as "connected growers" from a CDST perspective (de Bot et al., 2007).

However, what makes up the language and content subsystems and where do we draw the boundaries around them? Addressing these challenging questions, Larsen-Freeman (2017) drew on insights from the behavior of autopoietic systems, which "construct themselves by generating the very boundary conditions necessary for the creation and maintenance of their *self*-organization" (Witherington, 2011, p. 79, emphasis in original). Along these lines, of the different language subsystems especially pertinent to the creation and complexity of written texts are the lexical and syntactic subsystems, both of which have been extensively examined in developmental research on how L1 children and L2 adults gradually come to grips with complex language to develop linguistic maturity. Such maturity typically features an increasing gradation of syntactic complexity from parataxis/coordination, hypotaxis/subordination to lexicalization/nominalization (Givon, 1985; Ortega, 2012), and the emergence of more variegated and sophisticated vocabulary (Laufer & Nation, 1995). Content or thought complexity, as Han and Kang (2018) maintain, is also multi-faceted, manifesting itself in conceptual breadth and depth. When controlled for length, a text can show greater breadth with a variety of different main ideas, or greater depth with a few intricately elaborated ideas, but perhaps not both depth and breadth due to a potential trade-off between the two.

In summary, viewed through the lens of a complex dynamic system, text complexity is therefore primarily a function of the interweaving interrelationship between language and content complexity, emerging out of the local assembly of syntactic and lexical devices as well as the underlying breadth and depth of conception achieved. Figure 1 outlines the *theoretical* and *observational* make-up of the construct, based on Bulté and Housen's (2012) recommendation for construct definition and operationalization. How the subsystems at the observational level are *operationalized* and *measured* in the literature is expounded upon in the following section.

FIGURE 1
Theoretical and observational conceptualization of text complexity



Language and Content Measures of Text Complexity

A large bulk of L1 and L2 literature has been devoted to uncovering reliable and sensitive measures for indexing language complexity, and to a much lesser extent, content complexity. Syntactic complexity, the most extensively researched dimension of language complexity, has been assessed through length-based measures, such as mean length of sentence, clause, or T-unit (Bardovi-Harlig, 1992; Hunt, 1966; Norris & Ortega, 2009; Ortega, 2012; Wolfe-Quintero et al., 1998). However, the interpretations of those units have remained much debated. For instance, Bardovi-Harlig (1992) critiqued that T-units discount incipient learners' knowledge of coordination, suggesting that a sentence analysis would better capture the complexity of conjunction use such as *and*, *but* and *or*. However, Hunt (1966) advocated for T-units, arguing that language development in grade-school children was largely achieved through clause lengthening rather than the addition of T-units to sentences. Length measures have also been criticized for revealing little about the types of complex structures embedded therein and are thus difficult to interpret (Bulté & Housen, 2012; Norris & Ortega, 2009).

Unlike length measures that get principally at global complexity, ratio-based measures can tap syntactic complexity at the coordination, subordination, and phrasal levels, which have been shown to feature in L2 written production of different proficiency levels (Ortega, 2012; Norris & Ortega, 2009). Specifically, it is claimed that coordination dominates the complexity of L2 beginners' written production, and can be measured at the sentential level (mean number of T units per sentence) and the clausal level (mean number of coordinate phrases per T-unit or clause). On the other hand, subordination, calculated by mean number of dependent clauses per T-unit or clause, appears more characteristic of intermediate or advanced L2 writers' production (Norris & Ortega, 2009). This observation, however, was challenged by Biber et al. (2011) who contrasted grammatical features of research articles with conservation and found that clausal subordination was more common in conversational discourse than in academic writing that features predominantly complex phrases. Likewise, Ortega (2012), basing her argument on

Systemic Functional Linguistics Theory, maintains that language complexity at the advanced proficiency "shifts away from subordination and relocates in the processes of nominalization and grammatical metaphor" encoded in complex phrases (p. 142). Arguably, therefore, the syntactic complexity of written texts is expected to exhibit itself in the relative proportion of coordination, subordination, and nominal phrases the writer deploys to convey his/her intended meaning for the intended grade levels or target audience.

Lexical complexity, the other major dimension of language complexity, has largely been assessed through lexical frequency profiling (Cobb, 2007) and ratio-based measures (Bulté & Housen, 2012; Lu, 2012; Wolfe-Quintero et al., 1998). VocabProfile, a web-based profiling tool, has been widely used to analyze texts into the K-1 band (i.e., the most frequent 1,000 words), the K-2 band, the academic word list, and off-list words. The tool can also perform a full-blown profiling of texts into K-1 through K-20 bands. The underlying assumption is that higher bands indicate lower frequency and less concreteness, hence possibly greater lexical complexity. The ratio-based measures, on the other hand, are based primarily on the proportion of function words versus content words or type versus token contained within a text to capture lexical variation, sophistication, and density (Lu, 2012; Wolfe-Quintero et al., 1998). *Lexical variation* is generally measured by type/token ratios (TTR) of all words, nouns or verbs; *lexical sophistication* is assessed by TTR of sophisticated nouns, verbs, adjectives or adverbs; and *lexical density* is indexed by lexical words/total words or lexical words/function words. Han and Kang (2018) assert that this last measure and its counterpart, function words/total words, tap respectively into breadth and depth of thought complexity.

Idea unit has also served as a viable unit of analysis for gauging thought or content complexity and the level of comprehension demonstrated in learners' written recalls (e.g., Carrell, 1985; Shin et al., 2016), though its definition and operationalization vary across studies. For instance, Carrell (1985) defined idea units as *grammatical units* consisting of single clauses, including adverbial and relative clauses, infinitival constructions, gerundives, normalized verb phrases, prepositional phrases, appositives, and the like. Johnson (1970), by contrast, interpreted idea units as *pausal units* indicated by a native speaker's natural pauses during oral reading. Ellis and Barkhuizen (2005) viewed an idea unit as a message segment composed of "a topic and comment separated from contiguous units syntactically and/or intonationally" (p. 154). They also differentiated between major idea units that convey essential content of the text and minor idea units that embellish the main ideas in greater detail. Of those different definitions of idea units, Ellis and Barkhuizen's version seems better suited to the present study because dividing texts into topic-comment units better captures the propositional or semantic meaning of a text (Shin et al., 2016), which can complement the linguistic measures of text complexity.

Three observations from the above review of the literature are noteworthy for the present study. To begin with, content and language measures are seldom applied simultaneously to the same text, be it written by native speakers or L2 learners. Ellis and Barkhuizen (2005), nonetheless, is an exception. Assessing two texts produced by L2 learners in terms of both language and content complexity, they found that although the texts appeared similar in their length and TTR measures, one of them emerged as conceptually more complex, suggesting the need to use complementary measures. Secondly, the measures are not created equal and their sensitivity may depend on the text types (written vs. spoken), genres (narrative vs. informational), or proficiency levels if L2 learners' production is the target of analysis. As such, a reliable measure of text complexity is contingent upon judicious application of multiple measures to the appropriate text. These measures can address not only *whether* one text is more or less complex than another but also *how* they differ in terms of complexity. It is when armed with such refined understanding of text complexity we are more likely to pin down the source of complexity that may cause learning difficulty for L2 learners. More importantly, instead of adopting a reductionist approach by fixating on either language or content measures, these measures should be interpreted as constituents of a complex system that jointly contribute to the overall text complexity.

With these observations in place, the present study aimed to examine the complexity of science texts, which constitute an indispensable part of secondary and tertiary curricula across the U.S. as well as in content-based immersion programs worldwide (Lyster & Ballinger, 2011). Such texts are notorious for their subject-specific content and jargon that may pose a great learning challenge for native speakers of English, and English language learners and minority students in particular (e.g., Arya et al., 2017). The study thus addressed the following research questions: (1) Do high school and college science texts differ in terms of language and content complexity? (2) If so, how do they differ in these two dimensions?

METHOD

Text Selection

To address these research questions, science texts from both high school and college textbooks on four major hard science subjects, namely astronomy, biology, chemistry, and physics were sampled, resulting in four pairs of texts (available upon request). The rationale for selecting texts from both high school and college textbooks was that such texts were more likely to vary in complexity and could be controlled for genre variation, which may affect text complexity (Sheehan et al., 2010). However, no prior assumption was made that high school texts were less complex than college texts in terms of both language and content. The length of the texts was controlled at about 500 words so that complexity was not a function of length variation (Mesmer et al., 2012) and because such length stood a better chance in exemplifying coherent discourse use than texts of 200-300 words typically employed in extant studies (e.g., Crossley et al., 2008). Furthermore, to ensure that target texts were self-contained, they were sampled from the beginning of a chapter or section.

Automated Analysis of Language Complexity

The texts were analyzed in terms of syntactic complexity and lexical complexity. Lu's (2010) Syntactic Complexity Analyzer (SCA) was adopted to assess both the overall complexity and local syntactic complexity indexed by the amount of coordination, subordination, and phrasal complexity. Table 1 presents descriptions of all the units as defined in SCA, along with an example for each unit; Table 2 summarizes all the measures, along with their abbreviations. Although the tool was designed for analyzing advanced L2 learners' writing, it was deemed appropriate in this case because the syntactic parser employed was actually trained on native language data (Lu, 2010). In addition, Polio and Yoon (2018) showed considerably high reliability of SCA measures with manual coding of advanced L2 learners' writing except for the coordination measure of T-unit per sentence, whose reliability dipped to 0.75 due to run-on sentences. This exception, however, was less likely to be an issue in this study because the data analyzed constituted authentic texts from published textbooks.

TABLE 1
Descriptions of units of analysis and examples

Unit	Description	Example
Word	A single distinct meaningful unit of language bounded by space	*Cosmology*
Sentence	A group of words delimited by punctuation marks, e.g., ., *?*, *!*, and *""*.	*Every chemical reaction proceeds at its own rate.*
Clause	A structure with a subject and a finite verb	*A log, however, burns more slowly*.
Dependent clause	A finite adjective, adverbial or nominal clause	*When the starter flame is removed*, the reaction continues.
T-unit	An independent clause plus any dependent clauses	*When the starter flame is removed, the reaction continues.*
Complex T-unit	A T-unit that includes a dependent clause	*When the starter flame is removed*, *the reaction continues*.
Coordinate phrase	Adjective, adverb, noun, and verb phrases linked by a coordinating conjunction (e.g., *and*, *but*, *for*, *nor*, *or*, *so* and *yet*)	The motion of electrons is what we call *electricity or electric current*.
Complex nominal	a) Nouns with modifiers such as adjective, possessive, prepositional phrase, relative clause, participle, or appositive; b) Nominal clauses; c) Gerunds and infinitives functioning as subjects	a) *A whole skeletal muscle* is *an organ of the muscular system*. b) *Two electrons that are near each other* can interact. c) *Raising the temperature* speeds up reactions.
Verb phrase	A finite or non-finite verb phrase dominated by a clausal marker	Increased collision frequency *leads to a higher reaction rate*.

TABLE 2
Measures of syntactic complexity

Dimension	Measure	Abbreviation
Overall complexity	Mean length of sentence	MLS
	Mean length of T-unit	MLT
	Mean length of clause	MLC
Coordination	Coordinate phrases per T-unit	CP/T
	Coordinate phrases per clause	CP/C
	T-unit per sentence	T/S
Subordination	Clause per T-unit	C/T
	Complex T-unit per T-unit	CT/T
	Dependent clause per clause	DC/C

	Dependent clause per T-unit	DC/T
Phrasal complexity	Complex nominals per clause	CN/C
	Complex nominals per T-unit	CN/T
	Verb phrases per T-unit	VP/T

Lexical complexity was assessed through Lexical Complexity Analyzer (LCA) (Lu, 2012) and VocabProfile (Cobb, 2007). LCA was used to measure lexical density (i.e., ratio of lexical words to total number of words), lexical variation (i.e., ratio of word types to number of total words or tokens; TTR), and lexical sophistication (i.e., ratio of sophisticated word types to total number of word types). Sophisticated word types referred to those that were beyond the first 2,000 most frequently used word types. VocabProfile was employed to profile the texts into the K-1 though K-20 frequency bands generated from the British National Corpus (Cobb, 2007). The reason for obtaining finer-grained frequency bands instead of the K-1, K-2, and academic lists, was because the science texts sampled may contain low frequency words that were discriminative of text complexity but were not captured by the K-1 and K-2 bands.

Coding and Analysis of Content Complexity

The texts were first coded into idea units (IU), which were subsequently categorized into discourse functions (DF) of topics (T) or levels of elaboration, with each of these levels of coding exemplified in Table 3 and explained below. Following Ellis and Barkhuizen's (2005) definition, an idea unit (IU) here consisted of a topic and a comment, which typically corresponded to the subject and predicate of a sentence. As straightforward as this operationalization seemed, there were cases where the coding turned out indeterminate. Therefore, the following guiding criteria were developed through an iterative analysis of all the borderline cases: (a) sentential conjunctions (e.g., *and*, *but*, *so*, *nor*, *yet*, and *for*) were coded as separate IUs whereas phrasal conjunctions were not; (b) sentences with logical connectors (e.g., *when*, *although*, *if*, *since*, *while*, *so that*, *as soon as*, etc.) were coded as separate IUs; (c) appositives, present or passive participle phrases did not count; (d) relative clauses with relative pronouns and seven or more content words in the relative and main clauses combined were treated as IUs whereas those with less than seven content words were not. The seven-word cut-off point was motivated by the seven-plus/minus-two rule on information processing (Miller, 1956), which serves as a rough guideline instead of an absolute cut-off. An example for each of those cases is provided in Table 3: IUs 1 and 8-9 are examples for (a); IUs 20, 21, and 22 for (b); IU 23 for (c); IUs 3, 4, 5, and 8 for (d). The idea units were then classified into topic introduction and levels of elaboration. Topics represented the breadth of content complexity and the levels of elaboration the depth. Direct elaboration on the main topics counted as first level of elaboration (E1) whereas elaboration on the content contained in E1 counted as second level of elaboration (E2). If there was further expansion on any of the details on E2, it was counted as third level of elaboration (E3). The average length of the IUs and the proportion of texts devoted to each level of elaboration were calculated. The coding was performed iteratively until a 100% consistency was reached. A flowchart was also created for each text to better visualize their conceptual breadth and depth (available upon request).

TABLE 3
Sample Coding of Idea Units and Discourse Functions

#	Idea Unit (IU)	DF
1	Cosmology is the scientific inquiry into the nature, history, development, and fate of the universe.	T
2	By making assumptions […], scientists build models, or theories […] Mean length of clause	E1
3	That are not contradicted by the behavior of the observable universe	E2
4	That attempt to describe the universe as a whole, including its origins and its future.	E2
5	They use each model until something is found that contradicts it.	E1
6	Then the model must be modified or discarded.	E1
7	In 1905 Albert Einstein published his theory of special relativity	T
8	Which showed that space and time can be seen as aspects of a deeper structure, space-time	E1
9	And that mass and energy are really the same thing.	E1
	…	…
20	While it might be simpler to leave it out (by assigning it a value of zero),	E2
21	Einstein assigned it a positive value	E2
22	So that the universe would be essentially unchanging, as he expected.	E1
23	In 1929, however, U.S. astronomer Edwin Hubble announced an amazing discovery – evidence that the universe actually is expanding.	T
	…	…
28	To determine the distance to other galaxies, Hubble compared the brightness of certain giant stars in these galaxies to the brightness of presumably similar stars in our own galaxy,	E2
29	Whose distance had been calculated by a number of other, overlapping methods.	E3
30	To determine the speed at which a galaxy was receding from Earth, he observed its spectrum.	E2
31	Dark lines in the spectrum of colors can be identified as being produced by specific elements known on Earth.	E3
32	For these galaxies, the lines were shifted away from their normal wavelengths towards the red, long-wavelength part of the spectrum.	E3

RESULTS

Prior to reporting on the results for the language and content complexity of the texts, descriptive information on the texts is first presented in Table 4. The odd numbers (T1, T3, T5, and T7) represent texts sampled from high school textbooks on astronomy, biology, chemistry, and physics respectively, and the even numbers (T2, T4, T6, and T8) texts from corresponding college textbooks. As shown, each pair of texts was of similar length but of varying number of sentences, T-units, and clauses. High school texts generally contained more sentences, T-units, and clauses than college texts across the subjects, barring the two physics texts (T7 & T8) that had a reverse trend in the T-unit and clause measures. In addition, there were generally more clauses than T-units and sentences within each text, but the latter two did not differ drastically. This is unsurprising given that clauses were much granular units of analysis than sentences and T-units that only differed in the presence of sentential coordination. The texts were also subjected to three commonly used readability tools (Lexile Measure, Flesch Reading Ease, and

Flesch-Kincaid Grade Level) to estimate their overall complexity. The tools yielded consistent results that college texts appeared relatively more complex than high school texts, though the degree of difference between each set of texts varied. For example, the Lexile Measure revealed a much more substantial difference between the two chemistry texts (T5 & T6) than other texts, which was also born out in the Flesch Reading Ease and the Flesch Kincaid Grade Levels.

TABLE 4
Descriptive Statistics of the Texts and Readability Measures

Subject	Text	W	S	T	C	Lexile® Measure	Flesch Reading Ease	Flesch Kincaid Grade
Astronomy	T1	495	21	21	45	1200L - 1300L	48.21	11.09
	T2	496	17	21	41	1300L - 1400L	44.37	12.24
Biology	T3	500	30	31	42	1100L - 1200L	53.58	9.79
	T4	510	22	22	36	1300L - 1400L	46.42	12.54
Chemistry	T5	508	33	35	50	1000L - 1100L	48.44	10.32
	T6	508	22	25	44	1300L - 1400L	38.27	13.40
Physics	T7	495	27	29	46	1000L - 1100L	58.40	9.47
	T8	501	24	32	51	1100L - 1200L	39.20	12.98

Note. W = Word; S = Sentence; T = T-unit; C = Clause. Higher Flesch Reading Ease scores mean easier to read.

Table 5 summarizes results of syntactic complexity measures. As can be observed, college texts (T2, T4, T6, and T8) had on average much longer sentences (MLS) than high school texts (T1, T3, T5, and T7); however, minor differences were observed in the average length of T-units and clauses (MLT and MLC) and that T8 appeared shorter on both of these measures than T7. Comparing the texts across the measures of coordination, subordination, and phrasal complexity revealed that the college texts were consistently more complex in terms of CT/T, a measure of subornation, and CN/C and CN/T, two measures of phrasal complexity. No consistent pattern, however, was observed in the rest of the measures. Nonetheless, it is interesting to note that T6 scored higher than T5 on all the global and local measures across the board, which may have resulted from the greater complexity difference between those two texts as observed in the results of the readability tools.

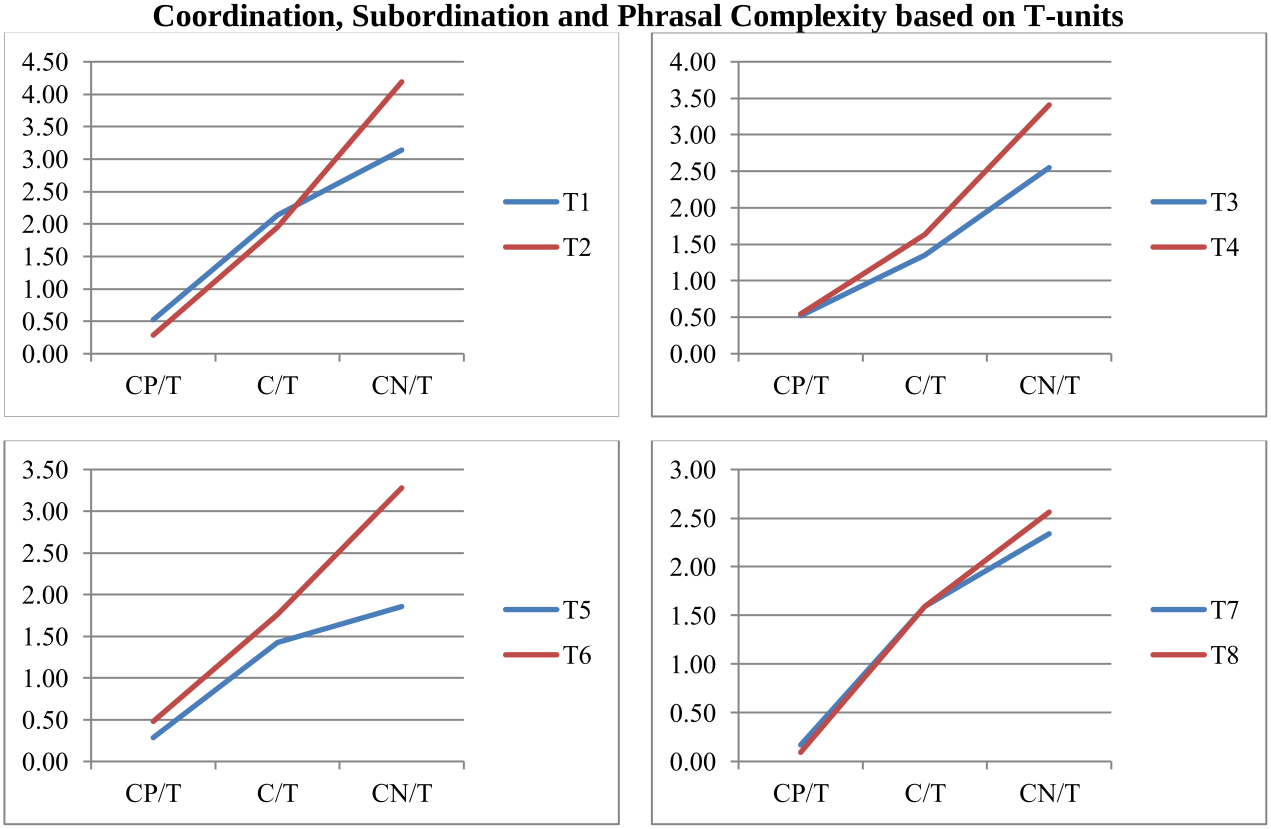
TABLE 5
Results of Syntactic Measures of Coordination, Subordination and Phrasal Complexity

	Overall complexity			**Coordination**			**Subordination**				**Phrasal complexity**		
Text	MLS	MLT	MLC	CP/T	CP/C	T/S	C/T	CT/T	DC/C	DC/T	CN/C	CN/T	VP/T
T1	23.57	23.57	11.00	0.52	0.24	1.00	2.14	0.67	0.53	1.14	1.47	3.14	2.52
T2	29.18	23.62	12.10	0.29	0.15	1.24	1.95	0.71	0.41	0.81	2.15	4.19	2.52
T3	16.67	16.13	11.90	0.52	0.38	1.03	1.35	0.32	0.29	0.39	1.88	2.55	1.94
T4	23.18	23.18	14.17	0.55	0.33	1.00	1.64	0.55	0.36	0.59	2.08	3.41	1.73
T5	15.39	14.51	10.16	0.29	0.20	1.06	1.43	0.37	0.30	0.43	1.30	1.86	1.86
T6	23.09	20.32	11.55	0.48	0.27	1.14	1.76	0.56	0.43	0.76	1.86	3.28	2.12
T7	18.33	17.07	10.76	0.17	0.11	1.07	1.59	0.52	0.37	0.59	1.48	2.34	2.10

T8	20.88	15.66	9.82	0.09	0.06	1.33	1.59	0.56	0.41	0.66	1.61	2.56	1.75

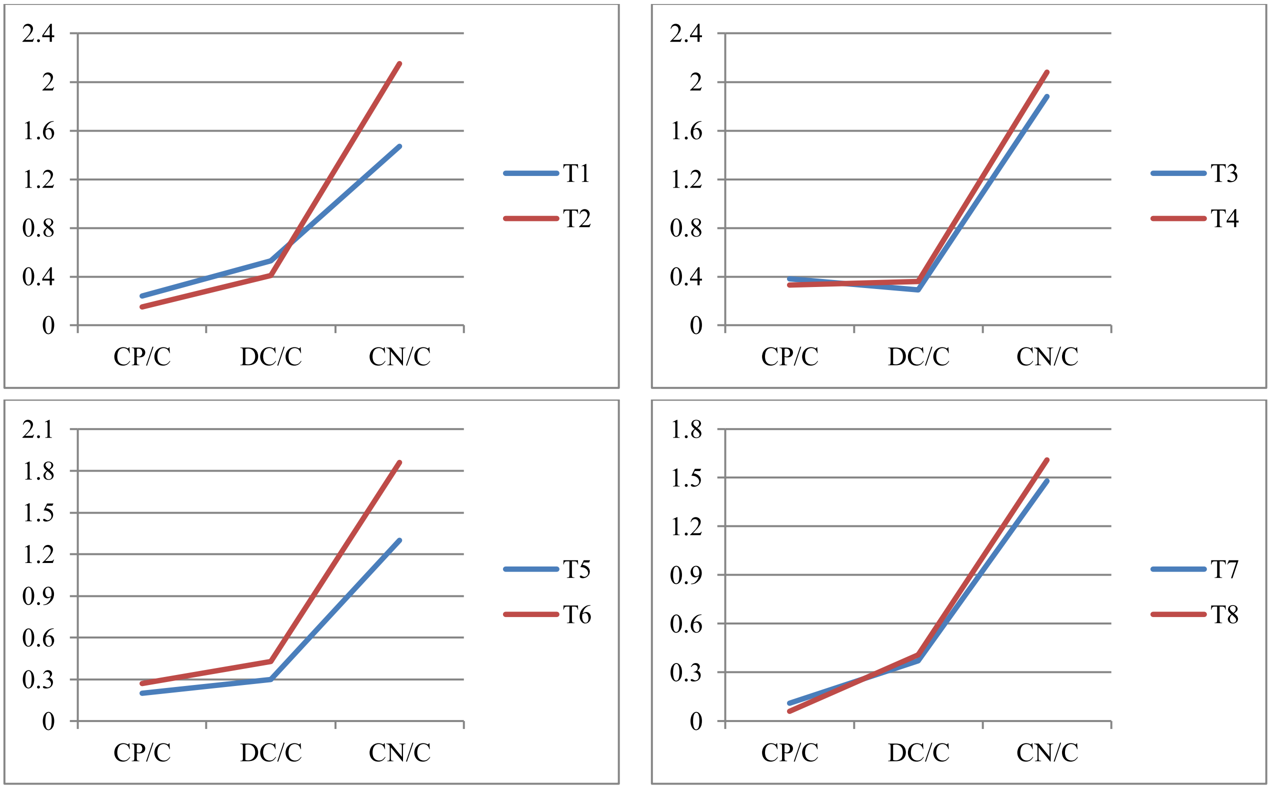
To provide a visual representation of the dynamic relationship between different local measures of coordination, subordination, and phrasal complexity within each text and across the four pairs of texts, both T-unit-based and clause-based measures were plotted in Figures 2 and 3. Both Figures 2 and 3 exhibited an overall increase from the amount of coordination, then subordination and finally to complex nominals between each pair of texts, although the increment from the amount of coordination to subordination was negligible in the clause-based measures shown in Figure 3. Moreover, neither the T-unit-based nor clause-based measures of coordination and subordination seemed to differentiate the four pairs of texts. However, the T-unit-based measures of phrasal complexity (CN/T and CN/C) in Figure 2 differentiated most texts except for T7 & T8 whereas its clause-base counterpart in Figure 3 failed to differentiate both T7 & T8 as well as T3 & T4, indicating that complex nominals, particularly when measured at the T-unit level, may serve as a more sensitive measure of syntactic complexity.

FIGURE 2
Coordination, Subordination and Phrasal Complexity based on T-units



Note. Blue = high school texts; red = college texts.

FIGURE 3
Coordination, Subordination and Phrasal Complexity based on Clauses



Note. Blue = high school texts; red = college texts.

In terms of lexical complexity, Table 6 summarizes results on lexical variation (type/token ratio), lexical diversity (content words/total number of words), and lexical sophistication (words that are beyond the K-1 and K-2 bands/total number of words). A close inspection of the results indicates that lexical variation was strikingly similar except for T1 & T2 and T7 & T8. In addition, aside from T7 & T8, high school texts seemed to have slightly higher lexical density whereas college texts exhibited consistently higher lexical sophistication than high school texts. Frequency bands from K-1 to K-20 levels were profiled for each text as well. Table 7 provides a cumulative breakdown of the results, with blanks indicating absence of vocabulary on those bands. K-16 through K-19 bands were excluded because one or no text fell on these bands. Also highlighted in the table were the 95% and 98% text coverage, two recommended vocabulary thresholds for learners to reach adequate comprehension of a text (Nation, 2006). The higher the threshold band, the more complex the vocabulary in the text. Overall, T1 & T2 and T7 & T8 would require K-4 level words to achieve 95% coverage, although T2 and T8 would entail K-12 and K-9 bands respectively to reach 98% coverage. By contrast, the lexical complexity of T3 & T4 and T5 & T6 were notably different, with the high school texts (T3 and T5) requiring K-8 and K-4 bands and the college texts (T4 and T6) requiring K-13 band to reach just 95% coverage. In conclusion, while the astronomy (T1 & T2) and physics texts (T7 & T8) were of similar lexical complexity at 95% coverage, the biology (T3 & T4) and chemistry (T5 & T6) texts required significantly different frequency bands.

TABLE 6
Results of Lexical Complexity Measures

Text	LV	LD	LS
T1	0.51	0.54	0.25
T2	0.45	0.53	0.26
T3	0.43	0.60	0.38
T4	0.43	0.56	0.41
T5	0.41	0.56	0.29
T6	0.40	0.53	0.31
T7	0.38	0.53	0.28
T8	0.41	0.59	0.37

Note. LV = Lexical variation; LD = Lexical density; LS = Lexical sophistication

TABLE 7
Frequency Profiles of K-1 Through K-20 Bands with 95% and 98% Coverage

K-band	T1	T2	T3	T4	T5	T6	T7	T8
K-1	73.64	73.24	61.60	59.92	66.27	65.29	70.97	63.69
K-2	82.90	85.11	78.20	75.83	84.70	78.11	83.89	74.80
K-3	92.56	92.35	88.00	85.46	92.15	86.99	93.23	91.47
K-4	95.98	95.57	90.80	89.00	96.27	88.96	99.00	94.64
K-5	96.58	95.97	92.00	90.96	97.45	91.52	99.99	96.23
K-6	96.98	96.17	93.00	91.55	98.82	94.28		97.02
K-7	97.99	96.97	93.60	92.73	99.21			97.22
K-8	98.39	97.17	95.80	93.32	99.41			97.42
K-9	98.59	97.37	96.00	93.91				97.82
K-10	98.79		96.40	94.30				
K-11		97.57		94.50				98.22
K-12		97.97	96.60	94.89	99.80	94.67		
K-13			96.80	95.28		95.85		
K-14			97.00	95.67		96.05		
K-15			97.40	95.87		96.44		
K-20			97.80					98.62
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Note. Total refers to the cumulative frequency of all the words contained in a text.

Regarding the content complexity, the number of idea units (IU) and the mean length of idea units (MLIU) are presented in Table 8. As shown, college texts (T2, T4, and T6) contained on average fewer but longer IUs than high school texts (T1, T3, and T5) except for the two physics texts (T7 and T8). Given roughly the same number of total words, longer IUs tend to encode denser information, thus indicating greater conceptual complexity. To illustrate, the following excerpts were taken from T3 and T4, two biology texts on the structure of skeletal muscle fibers. While the concepts of "*multinucleated*", "*many nuclei in a single cell*" and "*myoblasts*" were distributed over three IUs in T3, they were condensed into one IU in T4.

> *T3: Skeletal muscle fibers differ from "regular body cells (IU1). They are multinucleated, which means they have many nuclei in a single cell (IU2); during development many stem cells called myoblasts fuse together to form muscle fibers (IU3).*

T4: Skeletal muscle fibers are multinucleated cells formed by the fusion of numbers of elongated uninucleate cells called myoblasts (IU1).

TABLE 8
Number and Mean Length of Idea Units

Text	T1	T2	T3	T4	T5	T6	T7	T8
Word	495	496	500	510	508	508	495	501
IU	32	26	39	34	43	34	40	43
MLIU	15.47	19.08	12.82	15.00	11.81	14.94	12.38	11.65

Note. IU = Idea unit, MLIU = Mean length of idea unit

However, the idea units within the same texts were not of equal substance. Whereas some pertained to topic introduction, exhibiting conceptual breadth or width, others explicated the topic(s), indicating depth of conception. The proportion of texts serving each of these purposes is graphed in Figure 4. The blue bars represent the amount of text allocated for topic introduction, the red 1st level of elaboration, the green 2nd level of elaboration, and the purple 3rd level of elaboration. When controlled for length, a greater portion of the text designated for introducing different topics/subtopics or achieving conceptual breadth would result in a smaller portion for elaboration or developing conceptual depth, and vice versa. As an illustration, the flowcharts in Figure 5 outline the topics and levels of elaboration for T1 and T2. T1 touched upon the definition of cosmology, Einstein's special relativity, general relativity, and Hubble's discovery in chronological order, with each being expounded upon. T2, however, dealt with just one topic, a consensus on the contents of the universe, which was expanded to further illustrate the potential components of the content and the rationales for distinguishing unobservable dark matter from baryonic matter. The rationales were introduced after the mysterious dark energy in the original text, hence the dashed lines. Conceptually, T1 therefore demonstrated greater breadth and T2 greater depth, as revealed in the different proportions of the first two bars in Figure 4. Similarly, T3 covered the definition, structures, functions, and the interrelationship

FIGURE 4
Proportion of Topics and Levels of Elaboration for Texts 1-8

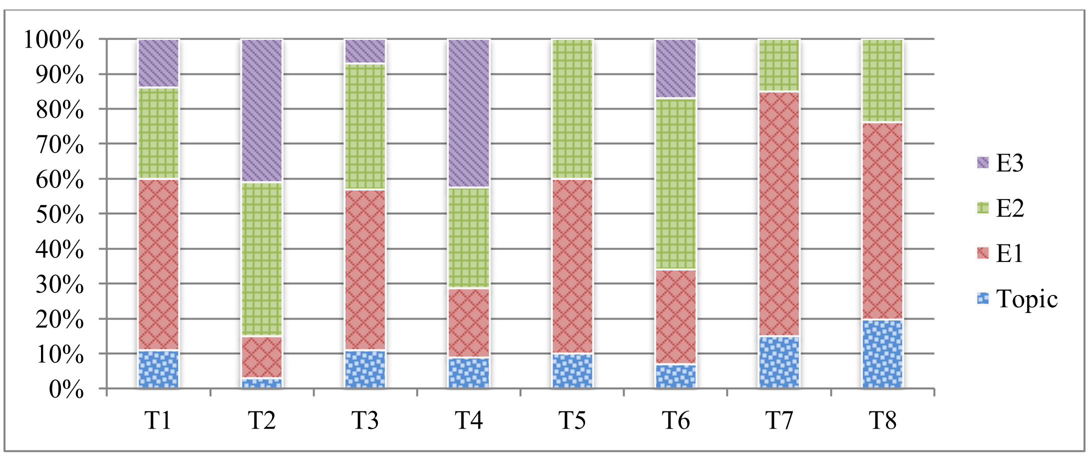


FIGURE 5
Flowcharts of Topics and Levels of Elaboration for T1 and T2

between skeletal muscle and skeletal muscle fibers whereas T4 zeroed in on elaborating the internal and external structures and functions of the latter. It seems that the expansion of ideas and the specificity of vocabulary entailed in T4 were highly contingent upon the inherent structure of skeletal muscle fibers, bounded by multiple layers of connective tissues from the *endomsium*, the *perimysium*, to the *epimysium*. With respect to the chemistry texts, T5 manifested greater breadth by introducing three factors (temperature, concentration, and particle size) that could be manipulated to alter the chemical reaction rates; on the contrary, T6 showed greater depth through elucidating the formation of a transition state and its effects on molecule reaction rates. Unlike the other three pairs of texts, T7 and T8 just had two levels of elaboration, as shown in Figure 4, with T8 exhibiting greater breath by covering a wider range of topics than T7. These findings, along with the results on the language complexity, are discussed in the section below.

DISCUSSION

Adopting a complex dynamic system perspective, this study attempted to bring together the language and content dimensions of text complexity, with each comprised of their respective subsystems functioning collectively to contribute to a text's overall complexity. By way of example, hard science texts from high school and college textbooks were sampled, compared and analyzed. The results showed that college texts in general exhibited greater language and content complexity than high school texts though there were variations in the degrees and ways of complexification across the four pairs of texts. In particular, college texts had on average longer sentences, more complex nominals and more sophisticated vocabulary than high school texts. Although no particular pattern on lexical variation was unearthed, college texts on the whole showed higher lexical sophistication but relatively lower lexical density than high school texts. In terms of content complexity, college texts had fewer topics, greater elaboration and longer idea units than high school texts overall, although the two physics texts (T7 & T8) showed similar depth but varying magnitudes of breadth.

Synthesizing the findings from the linguistic and conceptual dimensions of text complexity revealed a few patterns as well as a fair amount of variability both within and across the four pairs of texts. An interesting pattern that bridged the linguistic and conceptual complexity was that college texts in general exhibited greater lexical sophistication and conceptual depth, whereas high school texts showed greater lexical density and conceptual breadth. This finding from the lexical and content measures mirrors Han and Kang's (2018) proposal of using lexical density (ratio of lexical words to total number of words) to index breadth of thought. However, different from Han and Kang's proposal of using functional density[2] (ratio of function words to total number of words) to index depth of thought, this study found that lexical sophistication may indicate depth of thought or conceptual depth. Specifically, compared with high school texts, college texts across all four hard science subjects used relatively more low frequency lexical words to denote more complex concepts.

Another consistent pattern was that, of all the syntactic measures, college texts were found to have consistently much longer sentences than their corresponding high school texts, lending some credibility to the use of readability tools that are based primarily on sentence

[2] The reason why functional density was not calculated in the present study is because it can be inferred from lexical density and the ratios of functional and lexical density add up to 1.

length (Bernhardt, 1984). In addition, college texts featured considerably greater phrasal complexity, measured in terms of complex nominals over T-unit or clause (CN/T and CN/C), than high school texts. This finding corroborates Biber et al.'s (2011) observation that phrasal complexity is more characteristic of academic writing than clausal or subordinate complexity. More pertinently, complex nominals in the present study included nouns with modifiers such as attributive adjectives, prepositional phrases, appositives, gerunds, and infinitives. A defining attribute of such complex nominals lay in them being lexically dense and highly information-packed, which may have contributed to longer idea units and greater depth of elaboration in college texts than in high school texts, as illustrated in the two chemistry excerpts on in the results section.

Aside from these patterns, variability characterized both the magnitude of difference in complexity and the manner in which the texts differed. For example, T5 and T6 exhibited the most drastic differences across all the global and local syntactic measures whereas other texts differed chiefly in the number of complex nominals; T3 and T4 differed in T-unit-based complex nominals but not clause-based complex nominals; while most college texts demonstrated greater conceptual depth than high school texts, T7 and T8 varied principally in conceptual breadth, indicating a potential trade-off between those two dimensions. Such variability across the texts speaks to the necessity of employing multiple, fine-grained measures (Bulté & Housen, 2012; Norris & Ortega, 2009) to capture different dimensions of language and content complexity, and the importance of interpreting local measures of complexity in relation to each other and as meaningful wholes within a complex dynamic system.

In spite of the variability at the local measurement level and between texts, the patterns emerged from this study attested to the reciprocal relationship between language complexity and content complexity embedded within authentic texts. In other words, complex thoughts necessitate complex means to convey them (Givon, 1985; Verspoor, 2017). However, to make reading texts comprehensible for young readers or L2 learners, practitioners and textbook writers often resort to text simplification, typically achieved through reducing sentence length and complex structures, and replacing sophisticated vocabulary or expressions with more frequent ones (see DuBay, 2004, for a review). Empirical research examining text simplification (e.g., Yano, Long, & Ross, 1994) has revealed that although simplified texts may aid literal comprehension, they do not seem to promote acquisition of new linguistic structures largely because the linguistic richness and sophistication essential for language development might have unwittingly been stripped away. As Verspoor (2017) unerringly spells out, "simplicity comes at the expense of expressive power" (p. 147), indicating that linguistic simplification may unwittingly lead to the watering down of content complexity. It therefore follows that authentic texts, within which instances of the organic interplay between language complexity and content complexity are likely to manifest, should remain the primary focus of learning through reading.

CONCLUSION

This study conceived text complexity as a complex dynamic system, yielding from language complexity and content complexity and their respective subsystems. The results from comparing authentic hard science texts pointed to the trend that college texts were linguistically and conceptually more complex than high school texts. However, it should be noted that science texts may come with their unique discourse features, such as the prevalence of nominalization and phrasal expressions; therefore, it would be interesting to extend the approach to text complexity employed in this study to other genre types and topics that may manifest complexity in a different manner. Furthermore, although the focus of this study was on text complexity, which constitutes one significant source of processing difficulty, eliciting expert judgments or learner recall data in future research may help validate the operationalization and measurement of the construct, especially with regard to the breadth and depth of conceptual complexity. Previous research (e.g., Carrell, 1992) has shown that main ideas were easier to recall than elaborated details. Along this line, texts with greater conceptual depth would be expected to impose greater processing difficulty than those with greater conceptual breadth. This, however, is yet to be substantiated by empirical research. In spite of these limitations, this study represents a preliminary attempt towards a more theoretically motivated approach to, and a refined understanding of, text complexity, which hopefully will be taken up by future research to better assess and pinpoint potential learning difficulty resulting from the dynamic interplay of language complexity and content complexity.

REFERENCES

Arya, D. J., Hiebert, E. H., & Pearson, P. D. (2017). The effects of syntactic and lexical complexity on the comprehension of elementary science texts. *International Electronic Journal of Elementary Education*, *4*(1), 107–125.

Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, *26*(2), 390–395.

Bernhardt, E. B. (1984). Toward an information processing perspective in foreign language reading. *The Modern Language Journal*, *68*(4), 322–331.

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. (2012). Defining and 18perationalizing 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). Amsterdam/Philadelphia, PA: John Benjamins.

Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly*, *19*(4), 727–752.

Carrell, P. L. (1992). Awareness of text structure: Effects on recall. *Language Learning*, *42*(1), 1–18.

Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology, 11*, 38–63.

Common Core State Standards (CCSS) Initiative. (2010). *Common Core State Standards for English language arts and literacy in history/ social studies, science, and technical subjects*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Cromer, R. (1988). The cognition hypothesis revisited. In F. S. Kessel (Ed.), *The development of language and language researchers: Essays in honor of Roger Brown* (pp. 223–248). Hillsdale, NJ: Erlbaum.

Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, *42*(3), 475–493.

de Bot, K. (2017). Complexity theory and dynamic systems theory: Same or different? In L. Ortega & Z. Han (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman*, (pp. 51–58). Amsterdam: John Benjamins.

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.

DuBay, W.H. 2004. The principles of readability. Retrieved from http://www.impact-information.com/impactinfo/readability02.pdf.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford: Oxford University Press.

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.

Givon, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition* (Vol. 1, pp. 1008–1025). Hillsdale, NJ: Erlbaum.

Han, Z. H., & D'Angelo, A. (2009). Balancing between comprehension and acquisition: Proposing a dual approach. In Z-H. Han & N. J. Anderson (Eds.), *Second language reading research and instruction: Crossing the boundaries* (pp. 173–191). Michigan: The University of Michigan Press.

Han, Z.H., & Kang, E. Y. (2018). Revisiting the cognition hypothesis: Bridging a gap between the conceptual and the empirical. *International Journal of Applied Linguistics*, 1–15. Retrieved from https://doi.org/10.1111/ijal.12209.

Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English*, *43*(7), 732–739.

Johnson, R. E. (1970). Recall of prose as a function of the structural importance of the linguistic units. *Journal of Memory and Language*, *9*(1), 12.

Larsen-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, *18*(2), 141–165.

Larsen-Freeman, D. (2017). Complex theory: The lessons continue. In Z. H. Han & L. Ortega (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 11-50). Amsterdam: John Benjamins.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, *96*(2), 190–208.

Lyster, R., & Ballinger, S. (2011). Content-based language teaching: Convergent concerns across divergent contexts. *Language Teaching Research*, *15*(3), 279–288.

Masi, S. (2003). The literature on complexity. In L. Merlini Barbaresi (Ed.), *Complexity in Language and Text* (pp. 117–146). Pisa: PLUS.

Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading Research Quarterly*, *47*(3), 235–258.

Merlini Barbaresi, L. (2003). Towards a theory of text complexity. In L. Merlini Barbaresi (Ed.), *Complexity in Language and Text* (pp. 23–66). Pisa: PLUS.

Meyer, B. J. (1975). *The organization of prose and its effects on memory* (Vol. 1). Amsterdam, Netherlands: North-Holland Publishing.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63, 59–82

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. (2012). Interlanguage complexity: A construct in search of theoretical renewal. In B. Kortmann & B. Szmrecsanyi (Eds.), *Linguistic complexity: Second language acquisition, indigenization, contact* (pp. 127–155). Berlin/Boston, DE: De Gruyter.

Polio, C., & Yoon, H. J. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics*, *28*(1), 165–188.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics*, *22*(1), 27–57.

Robinson, P. (2003). The cognitive hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21 (2)*, 45–107.

Sharwood Smith, M. (1986). Comprehension versus acquisition: Two ways of processing input. *Applied Linguistics*, *7*(3), 239–256.

Sheehan, K. M., Kostin, I., Futagi, Y., & Flor, M. (2010). Generating automated text complexity classifications that are aligned with targeted text complexity standards. *ETS Research Report Series*, 2, i–44.

Shin, S. Y., Lidster, R., Sabraw, S., & Yeager, R. (2016). The effects of L2 proficiency differences in pairs on idea units in a collaborative text reconstruction task. *Language Teaching Research*, *20*(3), 366–386.

Valencia, S. W., Wixson, K. K., & Pearson, P. D. (2014). Putting text complexity in context: Refocusing on comprehension of complex text. *The Elementary School Journal*, *115*(2), 270–289.

Verspoor, M. (2017). Complex dynamic system theory and L2 pedagogy: Lessons to be learned. In Z. H. Han & L. Ortega (Eds.), *Complexity theory and language development: In celebration of Diane Larsen-Freeman* (pp. 143–162). Amsterdam: John Benjamins.

Witherington, D. C. (2011). Taking emergence seriously: The centrality of circular causality for dynamic systems approaches to development. *Human Development*, 54, 66–92.

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

Yano, Y., Long, M. H., & Ross, S. (1994). The effects of simplified and elaborated texts on foreign language reading comprehension. *Language Learning, 44*(2), 189–219.