

## LINGUISTIC COMPLEXITY IN ENGLISH TEXTBOOK AT SECONDARY LEVEL: A FUNCTIONAL GRAMMAR PERSPECTIVE

Majid Waseem  
[majidkb54@gmail.com](mailto:majidkb54@gmail.com)

MPhil Scholar, Department of English Language and Literature, The  
University of Lahore, Sargodha Campus, Pakistan

Ijaz Hussain  
[ijaz.hussain@ell.uol.edu.pk](mailto:ijaz.hussain@ell.uol.edu.pk)

Supervisor, Department of English Language and Literature, The  
University of Lahore, Sargodha Campus, Pakistan

### Abstract

The textbooks are the primary source of English input in Pakistani secondary classrooms and according to the systemic functional linguistics (SFL) theory, the language of well-graded materials is increasingly abstract, increasingly dense, and increasingly grammatically complex. The present study was an examination of the level of linguistic complexity in the revised Grade 9 English textbook published by the Punjab Curriculum and Textbook Board (PCTB) using three indicators of SFL namely lexical density (LD), nominalization and grammatical intricacy (GI). Eight prose units (6,229 running words, 648 ranking clauses) were annotated in UAM Corpus Tool 6 and each clause and nominalization candidate was checked manually using a corpus-based mixed analytical approach. The primary LD measure was Halliday's lexical items per ranking clause; the other measure used was Ure's percentage measure, used as a supplementary indicator. The level of complexity of the units varied. The range of LD was "Value of Time" from 4.03 to "Patriotism" at 7.29, of nominalization from 0.21 to 0.64, and of GI from 1.60 to 2.11, with the highest scores for "Patriotism" and "Quality Education: A Key to Success" and the lowest score for "Value of Time". The overall corpus figures for LD, nominalization and GI were 5.58, 0.46 and 1.85 respectively. All indicators were not monotonic across the series of units (Spearman  $\rho$  ranged from  $-.14$  to  $.31$  and all of them were greater than  $.45$ ). Both LD and nominalization increased together ( $\rho = .93$ ), while GI was independent of both such that there were no units that were always the simplest, or the most complex. The textbook thus has no systematic, sequential development of complexity. Sequencing of units, based upon an SFL-informed audit, and a clear teaching of nominalization and clause-combining are recommended.

**Keywords:** *Linguistic Complexity, Lexical Density, Nominalization, Grammatical Intricacy, Systemic Functional Linguistics, Textbook Analysis, Grade 9 English, Pakistan*

**Corresponding Author:** Majid Waseem (MPhil Scholar, Department of English Language and Literature, The University of Lahore, Sargodha Campus, Pakistan.)

**Email:** [majidkb54@gmail.com](mailto:majidkb54@gmail.com)

## 1. Introduction

Besides providing content for the lessons, textbooks also determine the language, level, and pace used in the classroom and teachers use textbooks to make pedagogical decisions (Putra & Lukmana, 2017). English textbooks written by the Punjab Curriculum and Textbook Board (PCTB) are compulsory in government and private secondary schools in Punjab and influence reading and writing and assessment in these schools. Grade 9 is a critical juncture in which students should shift from teacher-centered, concrete language of school to more abstract, academic language of schooling (Schleppegrell, 2004). Whether they are prepared for this transition, in part, is dependent on the language complexity level of their reading texts in the textbooks.

There are many different definitions of linguistic complexity, but SFL provides three well established indicators of text level (Halliday & Martin, 1993; Halliday & Matthiessen, 2014). Lexical density (LD) is a measure of the amount of lexical information (content information) in a text. A mode of grammatical metaphor is the nominalization, which takes a process or a quality and makes one noun, condensing and abstracting the information, and leading to the characteristic of academic writing. The grammatical intricacy (GI) refers to the extent to which meanings are expanded in clause chains in clause complexes. Written language tends to be denser and more nominal than spoken language, and complexity in learner texts is likely to grow gradually to allow learners to adjust to progressively more complex texts (Berendes et al., 2018). Sudden or erratic variation in these features can cause uneven cognitive demands in readers.

Books of several contexts have been studied, which reveal that this is not the case because this is not always found in the books, as reported by Ahmad (2022), Rahmah (2022), Bhatti & Bashir (2025, 2026). However, the updated Grade 9 English curriculum, which will be implemented in 2025, has not been studied using an SFL-informed, corpus-based approach that integrates LD, nominalization and GI throughout an entire book. This study addresses that lacuna. It answers four research questions

- **RQ1.** What is the level of lexical density in the reading texts of the revised Grade 9 English textbook?
- **RQ2.** How is nominalization used in these reading texts?
- **RQ3.** What degree of grammatical intricacy is present in these texts?
- **RQ4.** What do the three SFL features together reveal about the linguistic complexity of the textbook, and do they show a progressive pattern?

## 2. Literature Review

### 2.1. Lexical Density

Ure (1971) suggested that the lexical density of a text could be measured as the percentage of lexical items to the total number of running words and found that written texts would typically be over 40% and spoken and dialogic texts would be under 40%. Halliday (1989) proposed an alternative, the number of lexical items per ranking clause, which reflects the density of information in the clauses which carry meaning and is thus more favoured by SFL studies of written discourse. Lexical items are content words (nouns, main verbs, adjectives and manner or sentence adverbs), while grammatical items are function words (pronouns, determiners, prepositions, conjunctions and auxiliaries). There is evidence that higher density is associated with lower readability (Bakuuro, 2024), that it is associated with the abstract and information-rich language of upper-level materials (Solnyshkina et al., 2023), and that it can be compared using textbook studies (Fadhil et al., 2023).

## 2.2. Nominalization

Nominalization converts verbs (explore into exploration), adjectives (dense into density) and other nouns (leader into leadership) into nominal form, enabling the processes and qualities to be modified, countable and interrelated in the same way as things can be modified, counted and interrelated (Halliday & Martin, 1993). It is considered a crucial source of academic abstraction and a significant source of complexity in educational discourse (Cutrera et al., 2021). The existence of nominalisation in textbooks has already been documented (Jalilifar et al., 2014; Kaya & Apaydin, 2019; Fatima et al., 2024) and it is reported that nominalisation increases from grade to grade and verbal expression decreases (Solnyshkina et al., 2023).

## 2.3. Grammatical Intricacy

In SFL taxis links clauses together, with parataxis linking clauses of equal status and hypotaxis linking a dependent clause to a dominant one (Halliday & Matthiessen, 2014). The grammatical complexity is the number of ranking clauses/paratactic and hypotactic clauses (excluding embedded clauses) per clause complex. Spoken language tends to be more intricate and less lexical dense, whereas written language is more lexical dense and less intricate. So there are different ways of gaining complexity: in terms of packing information within clauses and in terms of chaining information across clauses.

## 2.4. Progression of Complexity in Textbooks

Berendes et al. (2018) suggest that as students get older, the difficulty of the school text they are required to access should increase. There have been mixed results to empirical testing of this expectation. Putra & Lukmana (2017); Mulyanti & Soeharto (2020) reported patterns of irregular complexities in Indonesian school books, while Rahmah (2022) indicated that the low complexity of school books did not show steady progress in four books in a series, and that the choice of word was the main factor influencing readability. Ahmad (2022) reported that the Intermediate English textbooks did not have a systematic and gradual development of linguistic complexity in Pakistan, while Bhatti and Bashir (2025, 2026) found that there was no development of syntactic complexity in any of the Intermediate English textbooks. The studies have been focused on comparing levels or on selected excerpts, but in this study these three levels of the SFL are determined on the prose units of one entire book.

## 3. Methodology

### 3.1. Design and Corpus

The study used a mixed analytical design, comprising of computational annotation and manual, theory-based verification. Ranking clauses and nominalizations are not automatically generated, instead must be interpreted in context and therefore the final data set should be regarded with that understanding. The corpus consisted of the 2025 edition of the revised Grade 9 English textbook published by the PCTB and posted on the official website of the board and was extracted in plain text with images, page numbers and formatting removed. Poems (Daffodils, If) and the dialogue play (Dialogue The Dear Departed) were omitted because the language of these pieces is condensed and generally finds rhythmic or conversational expression, which is not equivalent to expository or narrative prose. The final corpus consisted of eight prose units, namely, Savior of Mankind (U1), Patriotism (U2), Hazrat Asma (U3), Women Empowerment through Entrepreneurship (U4), Value of Time (U5), The Impact of Globalization on Culture and Economy (U6), Quality Education: A Key to

Success (U7), and The Silent Predator and Majestic Prey: Snow Leopard and Markhor (U8). The numbering of the units is sequential and follows the order as given in the textbook.

### 3.2. Analytical Framework and Measures

Three SFL indicators were computed for each unit:

- **Lexical density.** The main measure was Halliday (1989)'s lexical items to ranking clauses ratio. A supplementary descriptive measure was also calculated based on the percentage of lexical items (Ure, 1971), as words of this kind. Lexical items included all nouns (compound, but not compound with a hyphen), main verbs (in phrasal verbs), pure adjectives, and manner and sentence adverbs. Pronouns, determiners, modal, auxiliary and non-finite helping verbs, conjunctions, prepositions and adverbs of negation, degree, place and time were counted as grammatical items.
- **Nominalization.** Noun candidates (apart from proper nouns) were extracted using the part-of-speech tagging from UAM, filtered to exclude non-relevant concrete nouns, and only those that, in context, functioned as a process, action, event or quality considered as a noun, also as a grammatical metaphor, were kept. The types included verb-to-noun (plan → the plan; explore → exploration), adjective-to-noun (dense → density) and noun to noun (leader → leadership). Frequency is in terms of ranking clause.
- **Grammatical intricacy.** The number of clauses ranked divided by the number of clause complexes in the sentence was calculated as GI according to Halliday and Matthiessen (2014). The clauses were identified as either ranking (paratactic or hypotactic) or embedded, as the former distinction is not always reliably made in UAM.

### 3.3. Procedure and Sample Analyses

The text corpus was then entered into UAM Corpus Tool 6 (O'Donnell, 2008) for part of speech tagging, clause segmentation, transitivity annotation and frequency counts. Lexical items detected by the tool were segmented by the number of ranking clauses manually confirmed, GI was computed once the clauses were manually classified, and nominalizations were included in and were checked in the context. Pilot analyses of selected extracts from the text demonstrate the procedures: a 79-word extract from the textbook contained 40 lexical items in a total of five ranking clauses (Halliday LD = 8.0); and a sample of four clause complexes containing five ranking clauses gave GI = 1.25; a sample with nine nominalizations in five ranking clauses gave 1.8 nominalizations per ranking clause. No independent second coder was used and no inter rater reliability is reported as the agreement between automated and manual checks was used only as an internal consistency check in the pilot.

### 3.4. Data Analysis

Verified counts were used to get indices, per unit and for the corpus. Values reported to two decimals (one decimal for Ure's percentage). The direction of change from unit to unit was used descriptively as an indicator of progression, and as a supplementary check, Spearman rank correlations were used between unit order (1 to 8) and each index. Spearman correlations were also calculated between indices to describe the relationships between the three features. These correlations are descriptive and with 8 units a coefficient of .74 or greater is required for  $p < .05$ .

## 4. Results

### 4.1. Corpus Profile

Table 1 shows that the eight units had 6,229 running words, 3,619 lexical items, 351 clause complexes with 648 ranking clauses and 297 nominalizations. The range of size of the units varied

widely, from 344 words and 28 ranking clauses in U2, Patriotism, to 1,119 words (U4) and 118 ranking clauses (U7). The overall statistics for the corpus were 5.58 lexical items per ranking clause (58.1% according to Ure's measure), 0.46 instances of nominalisation per ranking clause and 1.85 instances of GI per ranking clause.

**Table 1.** *Corpus profile of the eight prose units of the revised Grade 9 English textbook*

|   | Unit                                                            | Words       | Lexical items | Ranking clauses | Clause complexes | Nominalizations |
|---|-----------------------------------------------------------------|-------------|---------------|-----------------|------------------|-----------------|
| 1 | Savior of Mankind                                               | 867         | 427           | 83              | 44               | 38              |
| 2 | Patriotism                                                      | 344         | 204           | 28              | 17               | 18              |
| 3 | Hazrat Asma                                                     | 615         | 315           | 64              | 40               | 19              |
| 4 | Women Empowerment through Entrepreneurship                      | 1119        | 691           | 110             | 58               | 59              |
| 5 | Value of Time                                                   | 663         | 347           | 86              | 46               | 18              |
| 6 | Impact of Globalization on Culture and Economy                  | 879         | 536           | 82              | 49               | 49              |
| 7 | Quality Education: A Key to Success                             | 972         | 659           | 118             | 56               | 62              |
| 8 | The Silent Predator and Majestic Prey: Snow Leopard and Markhor | 770         | 440           | 77              | 41               | 34              |
|   | <b>Total</b>                                                    | <b>6229</b> | <b>3619</b>   | <b>648</b>      | <b>351</b>       | <b>297</b>      |

*Note.* Units are listed in textbook order. Poems and the dialogue-based play were excluded.

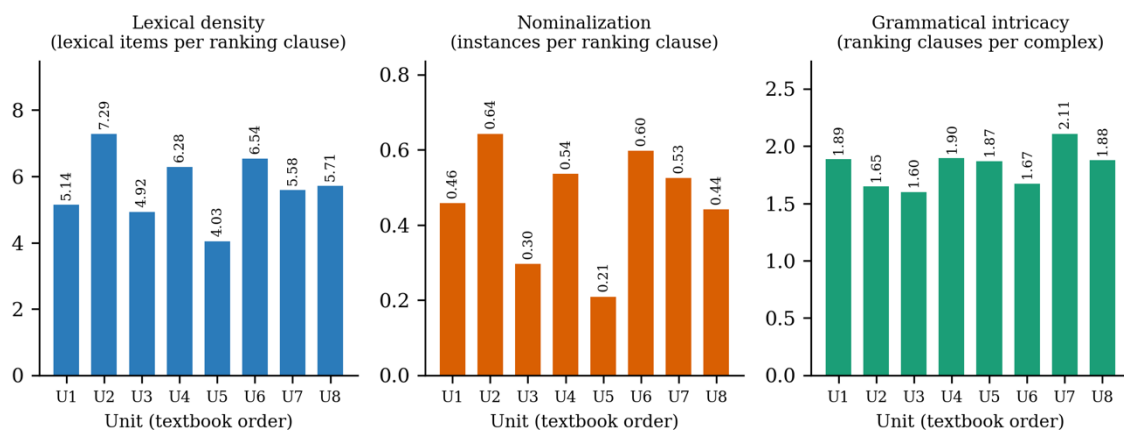
#### 4.2. Lexical Density (RQ1)

According to Halliday's main measure (Table 2, Figure 1), LD varied between 4.03 lexical items per ranking clause in U5 (Value of Time) and 7.29 in U2 (Patriotism) (1.8-fold difference). The densest units were U2 (7.29), expository/argumentative abstract vocabulary, compressed information; U6 (6.54) on globalization and U4 (6.28) on women's entrepreneurship. The least dense were U5 (4.03) and U3 (Hazrat Asma, 4.92), while U1 (5.14), U7 (5.58) and U8 (5.71) fell in between. The shorter the unit, the more clauses, with U2 being the densest (28 clauses) and longer units not systematically being denser. The value of LD was not raised throughout the book; it was raised and lowered across all seven transitions between units (rise, fall, rise, fall, rise, fall, rise) and not the lowest in the first unit, but in the fifth unit.

**Table 2.** *Lexical density by unit: Halliday's measure (principal) and Ure's measure (supplementary)*

|   | Unit                                                            | Lexical items | Ranking clauses | Halliday LD | Ure (%) | LD |
|---|-----------------------------------------------------------------|---------------|-----------------|-------------|---------|----|
| 1 | Savior of Mankind                                               | 427           | 83              | 5.14        | 49.3    |    |
| 2 | Patriotism                                                      | 204           | 28              | 7.29        | 59.3    |    |
| 3 | Hazrat Asma                                                     | 315           | 64              | 4.92        | 51.2    |    |
| 4 | Women Empowerment through Entrepreneurship                      | 691           | 110             | 6.28        | 61.8    |    |
| 5 | Value of Time                                                   | 347           | 86              | 4.03        | 52.3    |    |
| 6 | Impact of Globalization on Culture and Economy                  | 536           | 82              | 6.54        | 61.0    |    |
| 7 | Quality Education: A Key to Success                             | 659           | 118             | 5.58        | 67.8    |    |
| 8 | The Silent Predator and Majestic Prey: Snow Leopard and Markhor | 440           | 77              | 5.71        | 57.1    |    |

Note. Halliday LD = lexical items ÷ ranking clauses. Ure LD = lexical items ÷ total words × 100.

**Figure 1.** *Lexical density (Halliday), nominalization and grammatical intricacy across the eight units in textbook order (U1–U8).*

Ure's supplementary measure (Table 2, Figure 2) gave values from 49.3% (U1) to 67.8% (U7). Every unit exceeded the 40% level that Ure (1971) associated with written language, so all eight units show the lexical richness expected of written educational prose, and this measure also showed no gradual rise. The two measures ordered the units differently: Ure's measure ranked U7 highest,

whereas Halliday's measure placed it fifth, and U2 was highest by Halliday's measure but only fourth by Ure's (Spearman  $\rho$  between the two = .57,  $p = .139$ ). This difference supports treating the clause-based measure as the principal indicator, since it reflects the informational load carried by each clause and not merely the share of content words.

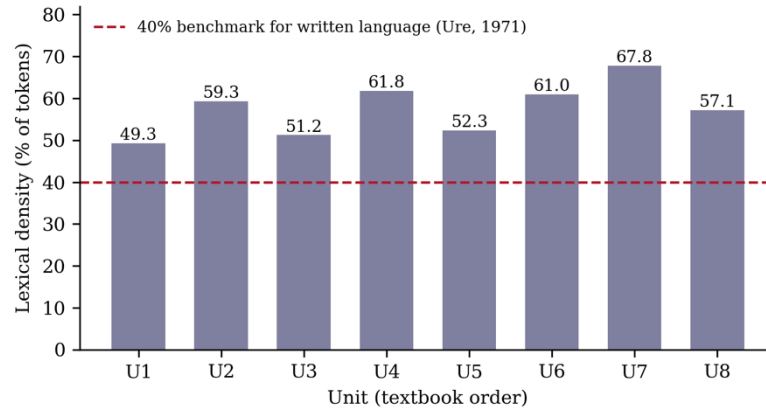


**Figure 2.** Lexical density by Ure's percentage measure (supplementary) across the eight units.

#### 4.3. Nominalization (RQ2)

Nominalization occurred in every unit but at very different rates (Table 3). The ratio ranged from 0.21 instances per ranking clause in U5 (Value of Time) to 0.64 in U2 (Patriotism), a range of 0.43, and the number of instances from 18 to 62. High ratios were found in U2 (0.64), U6 (0.60), U4 (0.54) and U7 (0.53), units that discuss abstract social, economic and educational concepts and compress ideas into nominal forms such as "innovation" and "entrepreneurship". Low ratios were found in U5 (0.21) and U3 (0.30), which are predominantly narrative or biographical and express meaning through processes and actions. U1 (0.46) and U8 (0.44) were moderate. As with LD, the ratio alternated from unit to unit (rise, fall, rise, fall, rise, fall, fall) rather than increasing.

|   | Unit                                           | Ranking clauses | Nominalizations | Per ranking clause |
|---|------------------------------------------------|-----------------|-----------------|--------------------|
| 1 | Savior of Mankind                              | 83              | 38              | 0.46               |
| 2 | Patriotism                                     | 28              | 18              | 0.64               |
| 3 | Hazrat Asma                                    | 64              | 19              | 0.30               |
| 4 | Women Empowerment through Entrepreneurship     | 110             | 59              | 0.54               |
| 5 | Value of Time                                  | 86              | 18              | 0.21               |
| 6 | Impact of Globalization on Culture and Economy | 82              | 49              | 0.60               |

|   |                                                                 |     |    |      |
|---|-----------------------------------------------------------------|-----|----|------|
| 7 | Quality Education: A Key to Success                             | 118 | 62 | 0.53 |
| 8 | The Silent Predator and Majestic Prey: Snow Leopard and Markhor | 77  | 34 | 0.44 |

**Table 3.** *Frequency and distribution of nominalization by unit*

*Note.* Instances were identified with UAM part-of-speech output and verified manually against SFL criteria for grammatical metaphor.

#### 4.4. Grammatical Intricacy (RQ3)

GI values ranged from 1.60 in U3 (Hazrat Asma) to 2.11 in U7 (Quality Education: A Key to Success), a range of 0.51 (Table 4). Two clusters can be distinguished. U7 (2.11), U4 (1.90), U1 (1.89), U8 (1.88) and U5 (1.87) had relatively high intricacy, meaning that meanings are elaborated through longer clause chains. U6 (1.67), U2 (1.65) and U3 (1.60) had lower intricacy and relied more on simple clauses and shorter complexes. All units, however, averaged more than 1.5 ranking clauses per complex, indicating moderate to relatively high clause complexity throughout. GI again showed no gradual increase; it fell between U1 and U3, rose in U4, fell to U6, and peaked in U7 before dropping in the final unit.

**Table 4.** *Grammatical intricacy by unit*

|   | Unit                                                            | Ranking clauses | Clause complexes | Grammatical intricacy |
|---|-----------------------------------------------------------------|-----------------|------------------|-----------------------|
| 1 | Savior of Mankind                                               | 83              | 44               | 1.89                  |
| 2 | Patriotism                                                      | 28              | 17               | 1.65                  |
| 3 | Hazrat Asma                                                     | 64              | 40               | 1.60                  |
| 4 | Women Empowerment through Entrepreneurship                      | 110             | 58               | 1.90                  |
| 5 | Value of Time                                                   | 86              | 46               | 1.87                  |
| 6 | Impact of Globalization on Culture and Economy                  | 82              | 49               | 1.67                  |
| 7 | Quality Education: A Key to Success                             | 118             | 56               | 2.11                  |
| 8 | The Silent Predator and Majestic Prey: Snow Leopard and Markhor | 77              | 41               | 1.88                  |

*Note.* GI = ranking clauses ÷ clause complexes.

#### 4.5. Integrated Profile and Progression (RQ4)

Table 5 and Figure 3 bring the three indicators together. The features did not move in the same direction. U2 (Patriotism) was the densest and most nominalized unit but had one of the lowest GI

values (1.65). U7 (Quality Education) had the highest GI (2.11) with mid-ranking LD (5.58) and fairly high nominalization (0.53). U5 (Value of Time) had the lowest LD (4.03) and nominalization (0.21) but a relatively high GI (1.87). Only U3 (Hazrat Asma) was low on all three indicators (LD 4.92, nominalization 0.30, GI 1.60), and it is the third unit, not the first. The first unit (U1) was not the simplest, especially in GI, where it ranked third, and the final unit (U8) was not the most complex on any indicator; its predecessor U7 exceeded it in GI, and U2, U6 and U4 exceeded it in LD.

**Table 5.** *Integrated linguistic complexity profile of the eight units: value (rank, 1 = highest)*

|   | Unit                                                            | Lexical density | Nominalization | Grammatical intricacy |
|---|-----------------------------------------------------------------|-----------------|----------------|-----------------------|
| 1 | Savior of Mankind                                               | 5.14 (6)        | 0.46 (5)       | 1.89 (3)              |
| 2 | Patriotism                                                      | 7.29 (1)        | 0.64 (1)       | 1.65 (7)              |
| 3 | Hazrat Asma                                                     | 4.92 (7)        | 0.30 (7)       | 1.60 (8)              |
| 4 | Women Empowerment through Entrepreneurship                      | 6.28 (3)        | 0.54 (3)       | 1.90 (2)              |
| 5 | Value of Time                                                   | 4.03 (8)        | 0.21 (8)       | 1.87 (5)              |
| 6 | Impact of Globalization on Culture and Economy                  | 6.54 (2)        | 0.60 (2)       | 1.67 (6)              |
| 7 | Quality Education: A Key to Success                             | 5.58 (5)        | 0.53 (4)       | 2.11 (1)              |
| 8 | The Silent Predator and Majestic Prey: Snow Leopard and Markhor | 5.71 (4)        | 0.44 (6)       | 1.88 (4)              |

*Note.* Lexical density = Halliday's lexical items per ranking clause; nominalization = instances per ranking clause.

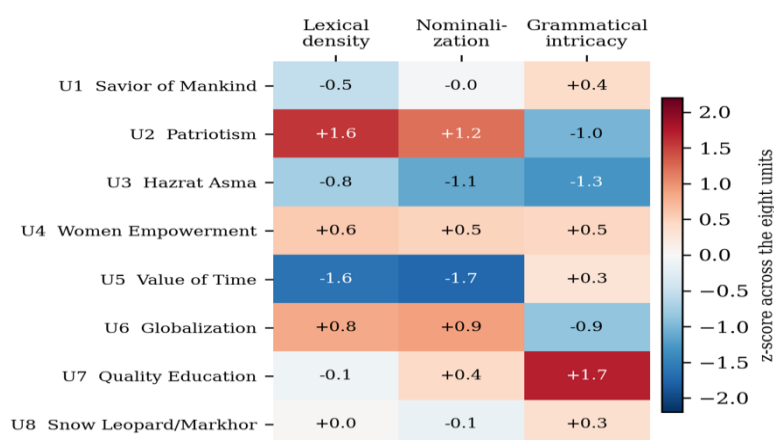


**Figure 3.** *Standardised (z-score) profile of the three indicators across the eight units; red = above the eight-unit mean, blue = below.*

Supplementary rank correlations confirmed the visual impression (Table 6). No indicator correlated significantly with the position of the unit in the book (LD  $\rho = .05$ ; nominalization  $\rho = -.14$ ; GI  $\rho = .31$ ; Ure LD  $\rho = .52$ ; all  $p > .18$ ), although with eight units only a strong monotonic trend could have been detected. Among the indicators, LD and nominalization were strongly and positively related ( $\rho = .93$ ,  $p = .001$ ; Figure 4), while GI was unrelated to either LD ( $\rho = -.05$ ,  $p = .911$ ) or nominalization ( $\rho = .05$ ,  $p = .911$ ). The LD–nominalization association is partly expected, since nominalizations are themselves lexical items and both indices share the same denominator.

**Table 6.** Descriptive statistics and rank correlation with unit order ( $n = 8$  units)

| Indicator                  | M    | SD   | Min  | Max  | CV (%) | $\rho$ with unit order (p) |
|----------------------------|------|------|------|------|--------|----------------------------|
| Lexical density (Halliday) | 5.69 | 1.02 | 4.03 | 7.29 | 17.9   | .05 (.911)                 |
| Nominalization per clause  | 0.46 | 0.15 | 0.21 | 0.64 | 31.8   | -.14 (.736)                |
| Grammatical intricacy      | 1.82 | 0.17 | 1.60 | 2.11 | 9.2    | .31 (.456)                 |
| Lexical density (Ure, %)   | 57.5 | 6.3  | 49.3 | 67.8 | 10.9   | .52 (.183)                 |

*Note.* M and SD are unweighted means and standard deviations across the eight units. CV = coefficient of variation. Correlations are Spearman rank correlations and are descriptive given  $n = 8$ .

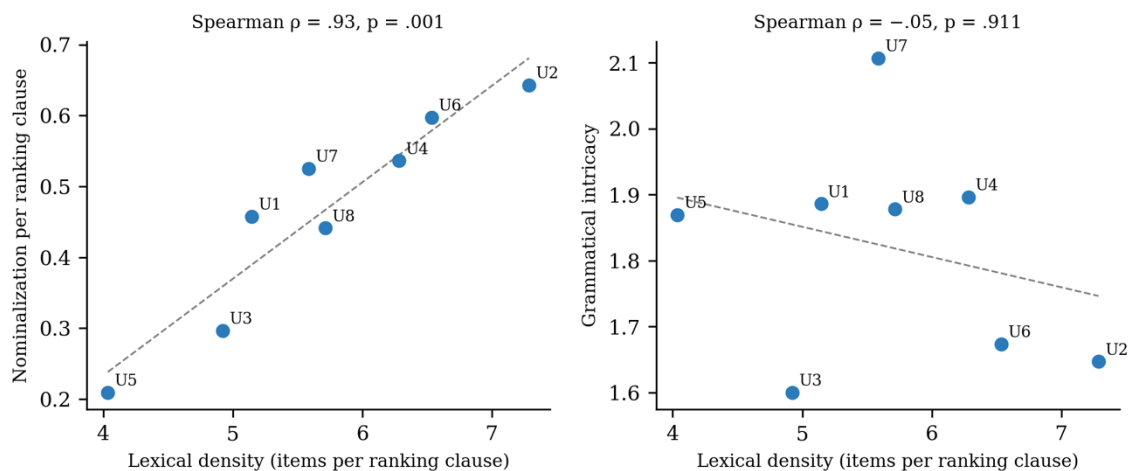


**Figure 4.** Relationships between lexical density and nominalization (left) and between lexical density and grammatical intricacy (right); labels are unit numbers.

## 5. Discussion

The central finding is that the revised Grade 9 textbook does not present a graded progression of linguistic complexity across its prose units. Lexical density, nominalization and grammatical intricacy each fluctuated from unit to unit, the simplest unit was the third rather than the first, and the final unit was neither the densest, the most nominalized nor the most intricate. This departs from the expectation that materials should increase in complexity gradually (Berendes et al., 2018) and is in line with reports of irregular complexity patterns in other textbooks (Putra & Lukmana, 2017; Mulyanti & Soeharto, 2020; Rahmah, 2022) and in Pakistani textbooks in particular (Ahmad, 2022; Bhatti &

Bashir, 2025, 2026). At the same time, the textbook is not simple: all units exceeded Ure's 40% benchmark for written language, nominalization occurred in every unit at roughly one instance for every two ranking clauses overall, and clause complexes averaged 1.85 ranking clauses.

The pattern of variation points to text type and topic as likely influences. The densest and most nominalized units are expository or argumentative texts on patriotism, globalization, entrepreneurship and education, while the least dense and least nominalized are biographical or narrative texts. This reading is interpretive, because genre was not coded systematically in this study. It is nevertheless consistent with the strong association between LD and nominalization, which suggests that abstraction through nominal forms is the main route by which some units pack information into clauses. Grammatical intricacy behaved differently: it was independent of the other two indicators, so a unit could be dense and nominal but syntactically simple (U2), or moderately dense but highly intricate (U7). Complexity in this textbook is therefore multidimensional, and no single indicator classifies a unit as easy or difficult. The two LD measures also ordered units differently, which shows how conclusions about a book can depend on the measure chosen.

The uneven sequence matters pedagogically. Readers who move from a low-density unit such as Value of Time (4.03) to a dense, highly nominalized one such as Patriotism (7.29; 0.64) meet a sharp rise in informational load and abstraction without preparatory scaffolding, and a limited working memory makes abrupt shifts more demanding. Because teachers rely on textbooks for input and sequencing, they may need to supply graded support themselves. Variation is not undesirable in itself: different genres and communicative purposes legitimately call for different language. The concern is variation that is unplanned and unsupported. Note also that progression was examined here within a single book across unit order; studies that compare whole books across grade levels address a different, though related, question.

## 6. Conclusion and Recommendations

Through three SFL indicators, this study profiled the linguistic complexity of the eight prose units of the revised Grade 9 English textbook. Lexical density ranged from 4.03 to 7.29 lexical items per ranking clause, nominalization from 0.21 to 0.64 instances per ranking clause, and grammatical intricacy from 1.60 to 2.11. Complexity was distributed unevenly and non-sequentially, and the three features did not progress together. The textbook offers rich linguistic resources for secondary learners, but it does not show a coordinated, systematic progression in linguistic demand.

Recommendations are as follows.

- **Textbook writers and curriculum planners:** review the order of units so that high-density texts follow easier texts or preparatory activities; use LD, nominalization and GI together when judging a unit's difficulty; and conduct an SFL-informed linguistic audit before publication to distinguish deliberate genre-related variation from unplanned gaps in progression.
- **Teachers:** teach nominalization explicitly in high-nominalization units (identify nominalized expressions, recover the underlying process and compare with congruent forms), and scaffold high-GI units through sentence decomposition, clause identification, guided paraphrasing and gradual clause combining.
- **Policy makers:** consider including SFL-based linguistic criteria in textbook approval and revision processes.

### 6.1. Limitations and Directions for Future Research

The analysis covers one textbook, eight prose units and three SFL indicators; other dimensions such as cohesion, discourse organisation and thematic development were not examined, and the findings should not be generalised to other boards or grades. Genre was not coded, so genre effects cannot be separated from sequence effects, and with eight units the correlational checks have low power. The study is text-based and did not collect classroom, teacher or learner data on comprehension. Clause classification and nominalization decisions were verified manually by one analyst with no independent second coder, so no inter-rater reliability statistic is available. Future research could add a second coder and an agreement statistic, code genre explicitly, compare several grades or boards, and relate the indices to readability and to learners' comprehension.

### References

- Ahmad, S. (2022). *Assessing linguistic complexity in elementary and secondary English textbooks: An SFL perspective* [Master's thesis, National University of Modern Languages, Islamabad].
- Bakuuro, J. (2024). In the belly of text complexity: Unravelling the nexus between lexical density and readability. *Athens Journal of Philology*, 11(3), 255–274. <https://doi.org/10.30958/ajp.11-3-4>
- Berendes, K., Vajjala, S., Meurers, D., Bryant, D., Wagner, W., Chinkina, M., & Trautwein, U. (2018). Reading demands in secondary school: Does the linguistic complexity of textbooks increase with grade level and the academic orientation of the school track? *Journal of Educational Psychology*, 110(4), 518–543.
- Bhatti, A. M., & Bashir, A. (2025). Linguistic complexity in Pakistani textbooks. *Al-Aasar*, 2(4), 294–303.
- Bhatti, A. M., & Bashir, A. (2026). From clauses to cognition: A computational analysis of linguistic complexity in Intermediate Book 2. *Journal of Applied Linguistics and TESOL*, 9(1).
- Cutrer, G. E., Massa, M. B., & Stipcich, S. (2021). Nominalization in processes of condensation of meanings in the teaching of saturated solution: A case study in a secondary school. *Ciência & Educação*, 27, e21028. <https://doi.org/10.1590/1516-731320210028>
- Fadhil, A., Gunawan, W., & Wirza, Y. (2023). Lexical density in EFL Indonesian textbooks: A comparative analysis. *Journal of Applied Linguistics and Literacy*, 7(1).
- Fatima, S., Parveen, M., & Qasim, H. M. (2024). Nominalization in Pakistani English news articles: A systemic functional linguistic perspective. *Annals of Human and Social Sciences*, 5(4), 696–704. [https://doi.org/10.35484/ahss.2024\(5-IV\)65](https://doi.org/10.35484/ahss.2024(5-IV)65)
- Halliday, M. A. K. (1989). *Spoken and written language* (2nd ed.). Oxford University Press.
- Halliday, M. A. K., & Martin, J. R. (1993). *Writing science: Literacy and discursive power*. Falmer Press.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.
- Jalilifar, A., Alipour, M., & Parsa, S. (2014). Comparative study of nominalization in applied linguistics and biology books. *Journal of Research in Applied Linguistics*, 5(1), 24–43.
- Kaya, M., & Apaydin, Z. (2019). In primary school 3th grade science textbook: Nominalization as grammatical metaphor. *Journal of Computer and Education Research*, 7(13), 35–61. <https://doi.org/10.18009/jcer.481884>
- Mulyanti, W., & Soeharto, P. P. (2020). Text complexity in English textbooks for junior high school: A systemic functional perspective. In *Proceedings of the Twelfth Conference on Applied*

*Linguistics* (CONAPLIN 2019) (pp. 217–222). Atlantis Press.  
<https://doi.org/10.2991/assehr.k.200406.044>

O'Donnell, M. (2008). The UAM CorpusTool: Software for corpus annotation and exploration. In *Proceedings of the XXVI Congreso de AESLA*.

Punjab Curriculum and Textbook Board. (2025). *English: Class IX* (Revised ed.). PCTB.

Putra, D. A., & Lukmana, I. (2017). Text complexity in senior high school English textbooks: A systemic functional perspective. *Indonesian Journal of Applied Linguistics*, 7(2), 436–444.  
<https://doi.org/10.17509/ijal.v7i2.8352>

Rahmah, N. F. (2022). The readability of junior's level in Mentari's textbooks: A systemic functional linguistics study. *JEPAL: Journal of English Pedagogy and Applied Linguistics*, 3(1).  
<https://doi.org/10.32627/jepal.v3i1.543>

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Lawrence Erlbaum.

Solnyshkina, M. I., Gatiyatullina, G. M., Kupriyanov, R. V., & Ziganshina, C. R. (2023). Lexical density as a complexity predictor: The case of science and social studies textbooks. *Research Result. Theoretical and Applied Linguistics*, 9(1). <https://doi.org/10.18413/2313-8912-2023-9-1-0-2>

Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443–452). Cambridge University Press.