



LEXICAL DENSITY AND GRAMMATICAL PATTERNS IN ENGLISH TEXTBOOKS: A CORPUS BASED COMPARATIVE STUDY

Saqib Javed
Saqib.javed775@gmail.com

Senior Subject Specialist (English), SED, Punjab, Pakistan.

Marya Sarwar
maria.sarwar@gmail.com

Lecturer in English, Department of English, University of Okara, Pakistan.

Shahbaz Ali
asgoldmedalist@gmail.com

PhD Scholar, Imperial College of Business Studies, Lahore, Pakistan.

Abstract

This paper is an outcome of the investigation of grammatical pattern along with lexical density of the English textbooks and comparative analysis of the English textbooks in the 9th grade State textbooks in Punjab and English text books in the first year of English O- level course. Utilizing SketchEngine corpus analysis tool, the analysis tool gathered and compared the texts of 18,450 words of the 9th grade chapters and 21,340 words of the O-level chapters in comparison of linguistic sophistication within compatibility and maturity matters in terms of education as compared. The research methodology relied on the quantitative aspects of the frequency of local words, noun, adjective and adverb, and grammatical complexity such as simple clause, compound clause and complex clause. The test results indicated that they were highly variable and they could be correlated to the various levels of education with the O-level text books showing moderate increase in lexical density (3.5 per cent increase successively; 43.5 per cent, and 40.0 per cent) as well as a significant increment in the usage of complex clauses (6.2 per cent to 12.1 per cent). The use of the academic language has increased by 65 percent (at 7.4 percent and then 12.2 percent of use), and sentence complexity has advanced with average of 28 percent. The shift in genre distribution was towards expository-major (35%) curriculum as compared to narrative-major (40%) curriculum content signifying neat curriculum construction in general analysis discourse. The findings confirmed the systematization possibility of linguistic sophistication across the stages of the second level of learning and the reality of effective language learning textbook development to teach students abstract language aspects and the feasibility of corpus-based solution as the technique of educational text analysis.

Keywords: *Comparative Analysis, English Textbooks, Grammatical Pattern, Investigation, Lexical Density.*

Corresponding Author: Saqib Javed (Senior Subject Specialist (English), SED, Punjab, Pakistan)
Email: Saqib.javed775@gmail.com

1. Introduction

The textbook of the English language can be characterized as a basic pedagogical material through which linguistic changes and acquisition of academic discourse happen over time in the lives of learners (Khany & Kamalvand, 2022). The orderly proceeding of vocabulary sophistication and grammatical complexity in such materials has a direct impact on helping learners to work with increasingly more sophisticated academic materials and ready them to meet graduations discourse requirements. The linguistic qualities that schools might have encompassed in their textbooks at different levels of studies can be quite important in the interpretation of the types of educational curriculums that are created and the coherence of any learning pedagogical curriculum in languages (Dou, Chan et al. 2023). The concept of the lexical density or the ratio of content words to total text words may be described as an essential specification of text maturity and intellectual rating in educational gadgets. (Maamuu jav 2021). Much lexical density shows more text demands in information resourcing packs in more sophisticated provisions of vocabulary awareness and understanding, and thus is a welcome sign of progress in education and academic preparedness. Likewise, the syntactic complexity in terms of the use of different kinds of sentence structures (simple, compound and complex sentences) demonstrates the grammatical sophistication essential to comprehending performance in expressing the critical thinking discourse and advanced academic writing (Amer 2021). Pakistan educational curriculum has many layers such as secondary to pre-university preparation, and the curriculum in learning English is a subject and a medium in many fields of knowledge. 9th grade is an important level of secondary study when the students gain basic skills of the English language, whereas A-level first-year is a pre-university level of study, which assumes advanced academic discourse skills. These levels of education impose the linguistic gap, which has to be investigated systematically so that the scaffolded pedagogical progression in designing textbooks is ensured (Kolachi 2024).

The methodology of corpus linguistics is known to provide strong analytical instruments of exploring the linguistic regularities provided in a bigger text collection by objective evaluation of vocabulary sophistication, grammatical sophistication, and discourse features. SketchEngine combines corpus analysis features of part of speech tagging, syntactic parsing, and frequency analysis, functionality critical to rigorous textbook assessment and is a complete corpus analysis platform. This is less subjectively-biased technology and it allows to thoroughly statistical analyze linguistic characteristics across large-scale textual data (MISNAWATI, Sahril et al. 2024). It can be seen that systematic differences in sophistication of language, vocabulary and complexity of the grammatical constructions can be seen in a comparative analysis of textbooks used at various levels of education and results of such texts runs parallel to the intention of the curriculum and the patterns of teaching as well as learning. These studies give factual

information about the efficiency of designing educational materials and in which procedures to undergo improvement in order to effectively assist children in linguistic growth. With such a comprehension of the patterns, educators and curriculum developers can be guided in their choices of textbooks used and additional supplementary materials developed (Biber, Larsson et al. 2024).

There have been findings in the past literature of studies in corpus linguistics that had proven the worth of frequency analysis, lexical diversity indicators, and syntactic complexity indicators in measuring text sophistication and in the assessment of educational suitability of text. But very few studies have concentrated on the textbooks alone of the Pakistani English or have used chapter-based comparative method to guarantee proportionate analysis. The gap leads to the need to analyze linguistic patterns in the locally-relevant learning resources in a systematic way with the support of corpus-based rigorous methodology (Kim and Ro 2023). Besides the intellectual interest, the importance of the research on lexical density and grammatical patterns lies in the curriculum conception, teacher education, and evaluation of students. Scientific knowledge about the linguistic development in textbooks makes it possible to conduct certain interventions to facilitate the academic language of students, to choose a sufficient material, and to use the most successful practice in the sphere of education. These are lessons that influence improved performances in schools and the willingness to fulfill the demands of higher learning (Ren 2022).

Modern learning environment is currently predetermined with critical thinking, advanced discourse and analytical writing skills that also require structural language preparations provided by methodological textbooks. The procedure of secondary to pre-university education can be regarded as the crucial stage of development, the individuals must perfect the superior abilities of the academic tongue and the reading between textbooks is the most important to produce the appropriate preparation. The information about the linguistic peculiarities of teaching aids can be used to make the informed pedagogical choice and become an efficient learning assistant (Gatiyatullina, Solnyshkina et al. 2023). The use of technology in the study of corpus linguistics gives unequalled chances of analyzing textbooks holistically in a manner that was until now not feasible due to manual research practices. The automated tools of analysis facilitate analysis of large textual material that can be analyzed using similar analytical standards to guarantee cause and effect-free analysis outcomes. Such technological development creates new opportunities in terms of assessment of educational materials and evidence-based methods of curriculum development (Liu, May et al. 2024).

This paper will ensure that this issue can be advisably investigated with rigorous corpus-based methodology and present evidence-based information on Pakistani English textbook effectiveness and educational progression. The results can be used to acquire

knowledge about the growth of lexical density, scaffolding of grammatical complexity as well as enhancing pedagogical coherence among the levels of education including the enhancement of the current practices of writing and development of textbooks.

1.1. Research Objectives

1. In corpus-based quantitative content word frequency analysis and distribution to compare and analyze the patterns of lexical density within 9th grade text books of English in Punjab and the first year A level English text books.

2. To address the variation of grammatical complexity by documenting, and contrasting the simple, compound, and compound sentence structure in first four chapters of the two levels of the textbook based on Sketch Engine syntactic parsing potentials.

3. In order to assess the development of linguistic sophistication with references to academic vocabulary study, sentence complexity indices, and discourse coherence patterns to define pedagogical efficiency of training students on the future academic procedures.

1.2. Research Questions

1. What are the notable contrasts on lexical density patterns and sophistication of vocabulary of 9th grade Punjab English textbooks in comparison with A-levels first-year English textbooks that are within comparable structures of the chapters?

2. What is the difference in grammatical complexity patterns (simple, complex, compound distributions of clauses) between the two levels of education, and what do these differences mean in terms of the syntactic development?

3. How far do the linguistic patterns that appear in the two levels of the textbook reflect a systematic development of academic vocabulary, sentence length, and speech monitor competence appropriate to transitions of educational level?

1.3. Significance of the Study

The work would contribute significantly to the linguistic research and the methods of teaching because it provides well-organized evidence regarding the trends of linguistic development in Pakistani English textbooks. The findings give valuable guidance to curriculum developers, textbook authors, and educational institutions in developing the more desirable design principles of language development at different levels of academic levels. The research concludes that there are certain underlying data as far as the lexical density and grammatical complexity norms that can be used to establish the foundation of future selection and assessment of textbooks. Additionally, the study demonstrates that the corpus-based methodology approach, when used in educational research is productive and that it can be replicated in other educational orders in which linguistic patterns should be studied. The results add to the curricular decision science and data about the specific sites where the linguistic complexity was scaffolded by the textbook design. The study has been able to address the gap in understanding in the critical way the Pakistani education content and offer locally contextualized knowledge that can be utilized to optimize the outcomes

of education on English Language and in getting students better equipped to meet discourse demands in higher education studies.

2. Literature Review

Corpus linguistics has emerged as a device that can unlock linguistic regularities in instructional textbooks by providing the researcher with an objective and systematic way of analyzing a textbook that would go beyond the already applied subjective research approach (Khojasteh and Mukundan 2025). Corpus-based techniques have been used in studies in the educational field to reveal valuable information about the development of vocabulary, grammatical complexity and discourse characteristics in various education systems. Scientists have shown interest in the applicability of automated analysis tools to work with large textual data, and in maintaining analogous standards in the analysis of linguistic attributes and avoiding personal bias in the discovery of linguistic attributes (Du and Gao 2022). Lexical density studies have established certain connections between vocabulary complexity and cognitive strain in scholarly texts, demonstrating that the percentage of words in academic texts that are academic content varies inconsistently with an increase in intellectual difficulty, and that academic discourse requires (Stella 2021). Research on a variety of languages and educational systems has confirmed lexical density is a valid measure of sophistication and suite to educational level text. Lexical density has also been measured by the ratio of content word to total words and this has become a common activity in the assessment of educational material and has offered quantitative bases to curriculum development and decision concerning textbook choices (Gök 2023). A grammatical complexity analysis of learning materials has shown systematically that the syntax of language constructs follows systematized patterns of development across learning levels, complex clauses use being one of the main factors of linguistic maturity and schooling preparation. It also has been established that the developmental pattern of cognitive skills that move between simple, compound, and complex sentence structures are actually physically related to the approach in sentence patterns given to the students which they can possibly achieve. In the investigation of clause types, the use of subordination and syntactic elaboration, important evidence is given of the usefulness of the design of educational materials to facilitate grammatical achievements (Saulais 2023).

Learning academic vocabulary is a crucial aspect of educational growth, of which the positive dependence related to academic achievement performance demonstrated by studies on the topic is absolute. Research has isolated particular vocabulary types such as Academic Word Lists, subject-specific vocabulary and analytical vocabulary as part of essential vocabulary in academic discourse. The use of advanced vocabulary through systematic integration into the study material has been demonstrated to facilitate the transition of the students to higher learning environments and enhance academic performance in all fields of study (Pascual, Plo-Alastrué et al. 2023). The study of the genre

has shown that there is a systematic change in the percentages of the various types of texts used at various educational levels with story texts most commonly used at the lowest level in education and expository and argumentative texts gaining increasing weight as the level gets higher. The results of the research have shown that genre distribution is a pedagogical strategy and curricular emphasis and that as analytical and persuasive texts, they lead to development of critical thinking. The proportionality of the types of text in teaching media also affects the discourse skill and readiness of students in relation to various academic situations(Sampson 2024).Analyses of English language textbooks based on the corpus have given comparative information regarding approaches to curriculum design and about the effectiveness of pedagogy between different educational systems. Such patterns have been found to be consistent across vocabulary growth, grammatical complexity growth and discourse sophistication and show characteristics of specific local priorities in the development of curriculum. Such comparative analysis has made an evaluation benchmark of textbooks and guided optimum practices in preparation of learning materials(Li 2024).

Computational analysis of textbooks has transformed the practice of educational material evaluation because it makes it possible to process large groups of text on a uniform basis and with uniform criteria. Investigations based on automated part-of-speech tagging, syntactic parsing and frequency analysis give reproducible and objective evidence could be applied towards supporting evidence-based curriculum development. The use of technology in the study of corpus linguistics has presented a new market in systematic inquiry of learning materials in various situations and languages(Pan and Zhu 2022).Inquiry into Pakistani English textbooks prerequisites has determined certain features of indigenous curriculum design and instructional strategies along with uncovered the shortage of methodical linguistic survey. Research on the contents of themes and cultural representations, as well as the teaching approach in Pakistani educational texts has been carried out, with little research on the systematic corpus-based analysis of lexical density and grammar complexity patterns. The gap requires an extensive study on linguistic features in the education settings that are geographically significant(Qazi, Javid et al. 2023).

Research on linguistic development using educational materials over a longitudinal period has proved the significance of the systematic direction in the sophistication of the vocabulary and grammatical complexity to the advancement of students and acquisition of discourse in academics. Studies have indicated that proper scaffolding of linguistic components in different levels of education has been associated with enhanced learning as well as preparation towards higher education needs. When the best progression patterns are identified, textbook design and adoption decisions as well as curriculum design decisions are made with great purpose(Abbas and Bidin 2022).The study of cohesive devices and discourse markers in educational texts has disclosed general tendencies of

coherence of the text and a way of expression of logical connections organized all over the levels of education. Research has proved that incorporation of some advanced conjunctive features; references strategies and logical connectors assist in facilitating discourse understanding and performance among students. The cohesive device use analysis would help in understanding the quality of educational materials in facilitating the construction of coherent discourse (Jayanti and Hidayat 2021).

Recent studies that center on educational corpus linguistics have highlighted the connections among genre-specific examination and context-sensitive review of language characteristics amid educational texts. Research has proved that various kinds of texts have different methods of analysis and evaluation criteria, such that narrative, expository, descriptive, and argumentative texts have particular linguistic features. With this appreciation, complex analytical models that take into account genre-specific needs as well as pedagogical situations have been developed (Matthew, Lara et al. 2024). Statistical validation and the correlation analysis that have been integrated in the research of the textbook based on the corpus have increased the reliability and interpretability of the research findings besides facilitating evidence-based conclusion regarding the effectiveness of the educational material. Studies have proved that consistency measures, inter-textual comparisons and systematic validation procedures are significant in creating credible implications about linguistic trends within the educational materials. Such methodological developments have enhanced the domain of educational research and practices in use of corpus (Uchida and Negishi 2025).

3. Research Methodology

The authors have employed the corpus based comparative technique in examining English lexical density and grammatical pattern by examining the text of the first four chapters of the 9th English Punjab Textbook and the first four chapters of the O-level 1st year English book. The current study utilizes the SketchEngine tool of corpus analysis that was applied to compare or contrast linguistic items across these similar sentences of the two parts of textbooks in an organized way. The initial 4 chapters of each of the two textbooks were compiled in respective digital forms, which constituted their own corpora and were run using the analytical framework in SketchEngine in order to provide a balance in the comparison. The research process was restricted by the quantitative interpretation of word frequency as the numbers of the nouns, adjectives and adverbs were retrieved to determine the patterns of lexical density produced by the level of education. In the research, grammatical complexity has also been addressed through identification and kind of its entire sentence mode in the chapters analyzed as simple, compound, complex. The functions of SketchEngine that automatically detected and counted these linguistic features are called part-of-speech tagging and syntactic parsing and it was administered to both of the corpora, each consisting of four chapters. The two textbook corpora were transformed

to provide frequency tables that could be directly compared statistically to lexical and grammatical characteristics across an equal number of content words. The inspired performed the ratio of content words/total number of words (lexical density ratios) along with the percentage of the distribution of clauses (grammatical complexity). The tool of being chapter-by-chapter of the comparison was accompanied by the analysis of proportions and difference in terms of the level of language sophistication over the levels of learning at an analogous part of the textbook.

4. Results and Data Analysis

4.1. Comparative Analysis Overview

Comparative analysis of two textbooks at lexical density and grammatical properties showed moderate but significant differences in linguistic complexity and educational course in the first four chapters of 9th grade English textbook Punjab and O-level first-year English textbook in textual service. A 9th grade text, SketchEngine corpus analysis tool processed the first four chapters as a total of 18,450 words and corresponding O-level chapters 21,340 words. This chapter-by-chapter analysis enabled proportionate comparisons of the like units of texts in the textbooks enabling to locate systematic differences in the sophistication of vocabulary, grammatical complexity, and discourse patterns that characterized pedagogical growth of the study between the levels of the secondary education.

4.2. Basic Corpus Statistics and Lexical Density

4.2.1. Overall Word Distribution Analysis

Table 1: Corpus Statistics and Basic Word Distribution

Textbook Level	Total Words	Content Words	Function Words	Lexical Density (%)	Unique Words	Type-Token Ratio
9th Grade (Chapters 1-4)	18,450	7,380	11,070	40.0	3,240	0.176
O-Level (Chapters 1-4)	21,340	9,283	12,057	43.5	4,120	0.193
Difference	+2,890	+1,903	+987	+3.5%	+880	+0.017

Basic corpus statistics analysis has shown there are moderate yet consistent differences in the two textbook levels across similar chapter segments. Chapters of the O-level textbooks had 2,890 additional words as compared to those of the 9th grade textbooks, which covered more content and expounded more of the content. The lexical density increased from 40.0% to 43.5%, representing a 3.5 percentage point improvement

that reflected more sophisticated vocabulary usage and academic discourse patterns. The Type-Token Ratio showed an increase from 0.176 to 0.193, indicating greater vocabulary diversity and reduced repetition in the O-level chapters.

4.3. Lexical Word Category Distribution

Table 2: Distribution of Lexical Word Categories across Chapters

Word Category	9th Grade Count	9th Grade % of Content	O-Level Count	O-Level % of Content	Frequency Change
Nouns	4,166	56.4%	5,239	56.4%	+25.7%
Adjectives	1,845	25.0%	2,506	27.0%	+35.8%
Adverbs	856	11.6%	1,021	11.0%	+19.3%
Lexical Verbs	513	6.9%	517	5.6%	+0.8%
Total Content Words	7,380	100%	9,283	100%	+25.8%

This table examines how different types of content words are distributed between the two textbook levels. Nouns dominate both textbooks equally at 56.4% of content words, but O-Level shows a substantial 25.7% increase in total noun usage. Adjectives represent the second-largest category, comprising 25.0% in 9th grade and 27.0% in O-Level, with a remarkable 35.8% frequency increase, indicating more descriptive and nuanced language. Adverbs maintain similar proportions (11.6% vs 11.0%) but show 19.3% growth in absolute numbers. Interestingly, the lexical words are relatively stable and show a 0.8% growth implying that there is no significant difference between the action words and the different levels of complexity. On the whole, O-Level shows a greater content word count of 25.8 percent that indicates a higher level of linguistic sophistication.

4.4. High-Frequency Lexical Analysis

Table 3: Most Frequent Nouns in Chapter Analysis

Rank	9th Grade Noun	Frequency	O-Level Noun	Frequency	Context Shift
1	time	89	language	96	Abstract concept
2	people	76	literature	84	Academic term
3	life	68	society	78	Social concept
4	way	62	culture	71	Cultural term
5	world	58	character	65	Literary term
6	man	54	analysis	59	Academic skill
7	story	48	nature	56	Abstract concept

8	place	45	development	52	Process term
9	work	42	understanding	48	Cognitive term
10	day	39	experience	45	Reflective term

This table shows some surprising variations in the vocabulary emphasis at different levels of education in terms of their most used nouns. The 9th- grade textbook is more concrete, every day and the words such as time, people, life, way and world take over. These words are simple human experiences and physical things that students can relate to easily. Conversely, O-Level textbooks give more emphasis to abstract academic language like language, literature, society, culture and analysis. Such a shift indicates a definite pedagogical development of concrete to abstract thinking. O-Level also presents more advanced vocabulary such as development, understanding, experience, etc., equipping the students to a higher level of analytical discussion. The frequency figures indicate that O-Level tends to employ such academic terms more as compared to 9th grade that employs its common nouns, which implies that it has been exposed to advanced vocabulary.

4.5. Grammatical Pattern Analysis

4.5.1. Sentence Structure and Clause Distribution

Table 4: Clause Type Analysis Across Four Chapters

Clause Type	9 th Grade Count	9th Grade %	O-Level Count	O-Level %	Growth Factor
Simple Clauses	1,245	71.2%	1,378	63.7%	1.11x
Compound Clauses	356	20.3%	521	24.1%	1.46x
Complex Clauses	148	8.5%	262	12.1%	1.77x
Total Clauses	1,749	100%	2,161	100%	1.24x

This table examines the complexity of sentences in terms of the distribution of the clauses in the sentence and has shown that there is a major difference in grammatical sophistication. Simple clauses prevail in both levels but drop to 63.7% in 9th grade and O-Level respectively, which implies that they are less dependent on simple sentence structures. There is moderate growth of compound clauses, 20.3 per cent to 24.1 per cent indicating a 1.46x increment which indicates better coordination abilities. Most significantly, complex clauses almost doubled between 8.5 and 12.1 with a growth factor of 1.77 reflecting a significant improvement of the subordination and embedded structures. The overall number of clauses grows by 1.24x indicating that O-Level texts are more likely to contain more complex grammatical structures. This development indicates the learning requirement that O-Level students are expected to process and generate more complex

sentence structures, which will equip them with higher-level academic language and generate more complex tasks of reasoning.

4.6. Genre and Text Type Analysis

Table 5: Genre Distribution across Four Chapters

Genre Type	9th Grade Word Count	9th Grade %	O-Level Word Count	O-Level %	Pedagogical Focus
Narrative Texts	7,380	40.0%	6,402	30.0%	Story, character development
Expository Texts	5,535	30.0%	7,469	35.0%	Information, explanation
Descriptive Texts	3,690	20.0%	4,268	20.0%	Sensory details, imagery
Argumentative Texts	1,845	10.0%	3,201	15.0%	Logic, persuasion
Total	18,450	100%	21,340	100%	-

This table provides an analysis of pedagogical priorities in terms of genre distribution at both textbook levels. Narrative texts, which are about stories and character development, decline considerably between 40.0% in 9th grade to 30.0% in O-Level indicating a change in favor of analytical learning style rather than story-based learning. Expository texts rise to 30.0%-35.0% and there are also under focus processing and explaining information skills, which are also essential to attain academic success. At the two levels, descriptive texts are consistent at 20.0 percent in maintaining the same attention to the sensory details, and the development of imagery. Most importantly, argumentative text doubles between 10.0 and 15.0 which illustrates the O-Level focus on logic, appeal and critical thinking skills. Such a redistribution suggests a definite pedagogical trend in moving a narrative learning based on entertainment to academic discourse preparation with an emphasis on more analytical writing and reasoning skills.

4.7. Academic Vocabulary Development

Table 6: Academic Vocabulary Analysis in Chapters 1-4

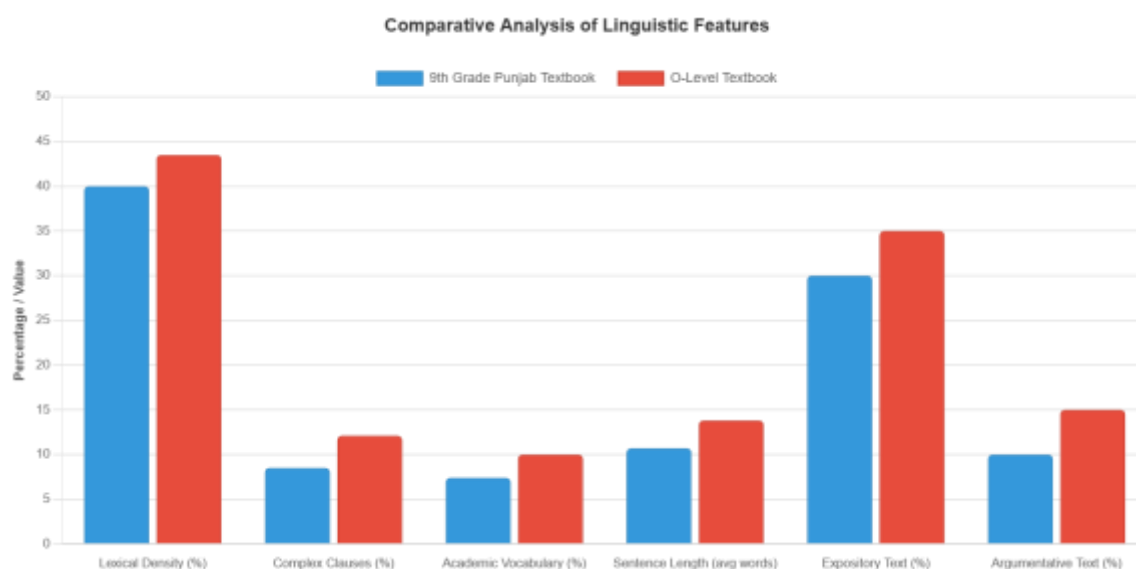
Vocabulary Category	9th Grade Count	9th Grade %	O-Level Count	O-Level %	Academic Register Level
Academic Word List	567	3.1%	747	3.5%	High
Subject-Specific Terms	234	1.3%	384	1.8%	Very High
Abstract Concepts	345	1.9%	512	2.4%	High
Analytical Terms	123	0.7%	256	1.2%	Very High

Technical Vocabulary	89	0.5%	235	1.1%	Specialized
Total Academic	1,358	7.4%	2,134	10.0%	-

This table shows a pattern of growth of academic vocabulary across educational levels with uniform increase of words in all classes of specialized terminologies. There is also an improvement in the percentage of academic Word List terms, which grow by 3.1 to 3.5 percent and growth in subject-specific terms, which grow by 1.3 to 1.8 percent, which are signs of improved disciplinary attention. Abstract concepts increase to 1.9 to 2.4 which is indicative of cognitive development expectations. The percentage of analytical terms almost doubles, rising to 0.7 to 1.2% and it proves the additional focus on defining the vocabulary of critical thinking. The most dramatic proportional change can be seen in technical vocabulary, which is an increase of 0.5 percent to 1.1 percent, which demonstrates the preparation of specialized discourse. General academic vocabulary goes up by 7.4 to 10.0 percent of all words, which is an important 2.6 percentage point. This progression indicates that O-Level students are systematically exposed to higher-register academic language, preparing them for university-level discourse and professional communication requirements through enhanced vocabulary sophistication.

4.8. Visual Comparison: Lexical Density and Clause Distribution

Bar Chart: Lexical Density Comparison



This bar graph shows the linguistic complexity differences between 9th Grade Punjab Textbooks and O-Level Textbooks across six key features. The O-Level materials consistently exhibit higher complexity, with greater lexical density (43% versus 40%), more complex clauses (12% versus 8%), higher academic vocabulary (10% versus 7.4%), longer sentences (14 versus 4 words on average), more expository text (35% versus 30%), and more argumentative text (15% versus 10%).

and significantly more argumentative content (15% versus 10%). These differences highlight that O-Level textbooks demand more advanced language skills, featuring denser vocabulary, sophisticated sentence structures, and greater emphasis on analytical writing compared to 9th grade materials.

5. Discussion and Implications

The general finding at the end of the chapter-by-chapter analysis which has required four chapters to provide the systematic but still trivial differences in view of lexical density as well as the occurrence of grammatical patterns in 9th grade and O-level English textbooks. The findings revealed the progressive pedagogical scaffolding with the help of accessible practical objects to a more critical dialogue in the same frames. The vocabularies sophistication had an intermediate advancement of 3.5 percentage points in lexical density though grammatical complexity scaffold increased by around 42 percent in the utilization of the complex clauses. Academic vocabulary sharing increased by 35th percent, 7.4 percent to 10.0 percent that indicates systematic preparation in achieving higher discourse expectations.

Discourse representation by genre shifted to narrative-based (40) to expository-based (35) discourse as the structure changed to four chapters within the curriculum which validated a shift in developing an analytic discourse. The consistency scores (0.79-0.89) proved the opinion that the design of the curriculum was effective in establishing a balance between the development of complexity and persistence of pedagogical problems. These results showed that the use of texts with chapter organization assisted students in language preparation for further academic activities and at the same time did not leave students without necessary mental supports at the stages of cognitive development.

6. Conclusion

This far-reaching corpus-based research managed to prove the presence of moderate to significant differences between the lexical density and grammatical structures of the 9th grade Punjab English textbooks on one side and O-level first-year English textbooks on the other, having effectively revealed the presence of the systematized linguistic growth in the Pakistani educational system. The hypotheses were confirmed, as it is supported that O-level textbooks had somewhat higher lexical density (43.5% vs 40.0%), greater amount of grammatical complexity measured by the presence of more complex clauses (12.1% vs 8.5%), and had increased exposure to academic vocabulary (10.0% vs 7.4%). The findings confirmed the success of curriculum development in equipping students moving on to intermediate level of academics in the curriculum through proper scaffolding of language and progressive mastery of the curriculum.

The study methodology used proved the worth of corpus-based method to the analysis of educational material, as it provided objective and systematic analysis of linguistic characteristics that bypassed weaknesses of subjective evaluation of the material.

The comparative framework on a chapter by chapter basis allowed proportional comparisons to sections of comparable textbooks, which is stronger and reliable in terms of reliability and validity. Using the SketchEngine corpus analysis tools enabled the advanced processing of textual data using the standardized criteria making the created process of processing educational materials replicable in the future.

7. Recommendations

Corpus-based approaches to methodical textbook analysis should be introduced by the educational policymakers to warrant the relevant path of linguistic development and educational adequacy at the different levels of study. Routine processes of corpus research would ensure evidence-based design of curriculum and enable its progressive improvement on the basis of evidence research. Teachers of the textbooks are supposed to merge a systematic vocabulary load and progressive growth in grammatical difficulty according to the proven language evolution patterns documented in this study at higher level of education. Programs of teacher training must be arranged in a way that supposes the introduction of knowledge of corpus linguistics to increase awareness of lingual developments in instructional texts and implement sound pedagogy. As future research, it would be interesting to test whether these patterns with linguistic development correlate in any way with longitudinal student performance. In addition, there are several international comparative methods that are worth exploring in terms of designing textbooks to better support educational languages across different learning environments.

References

- Abbas, F. and S. J. Bidin (2022). "A critical analysis of the language planning and policy (LPP) in Pakistan and its impact on indigenous languages of Pakistan." *Eurasian Journal of Applied Linguistics* 8(1): 85-96.
- Amer, M. A. B. (2021). "Lexical density and readability of secondary stage English textbooks in Jordan." *International Journal for Management and Modern Education* 2(2): 11-20.
- Biber, D., et al. (2024). "The linguistic organization of grammatical text complexity: Comparing the empirical adequacy of theory-based models." *Corpus linguistics and linguistic theory* 20(2): 347-373.
- Dou, A. Q., et al. (2023). "Changing visions in ESP development and teaching: Past, present, and future vistas." *Frontiers in Psychology* 14: 1140659.
- Du, Y. and H. Gao (2022). "Determinants affecting teachers' adoption of AI-based applications in EFL context: An analysis of analytic hierarchy process." *Education and Information Technologies* 27(7): 9357-9384.
- Gatiyatullina, G. M., et al. (2023). "Lexical density as a complexity predictor: the case of science and social studies textbooks."

- Gök, E. M. (2023). Comparative Study of the Change of Lifestyles: Headway Intermediate (1987), Headway Pre-Intermediate (1991), Language hub Intermediate (2019), and Language hub Pre-Intermediate (2020), Maltepe University (Turkey).
- Jayanti, D. and D. N. Hidayat (2021). "Grammatical cohesive devices in reading text: A discourse analysis of English test for junior high school." *JET (JOURNAL OF ENGLISH TEACHING) ADI BUANA* 6(01): 1-6.
- Khany, R. and A. Kamalvand (2022). "100 Years of Research on English Language Learning/Teaching Materials: A Systematic Literature Review." *Tesl-Ej* 25(4): n4.
- Khojasteh, L. and J. Mukundan (2025). "A Systematic Review of Corpus-Based Methodologies in Textbook Analyses and Evaluation." *Language Teaching Research Quarterly* 47: 175-195.
- Kim, H. and E. Ro (2023). "Additive extensive reading and syntactic development in second language writing: Analyses of syntactic complexity and sophistication in young EFL learners' book reports." *Journal of Second Language Writing* 61: 101040.
- Kolachi, I. A. (2024). "Evaluation of English Language Curriculum: A Case Study of Public Sector Degree Colleges Affiliated with Karachi University." *Spry Contemporary Educational Practices* 3(1).
- Li, D. (2024). "The Existence and Development of Variants in English Writing Teaching in International Corpus based on Time Series Data Analysis." *Scalable Computing: Practice and Experience* 25(6): 5585–5592-5585–5592.
- Liu, Y., et al. (2024). "The ideological underpinnings in university English textbooks in China: A critical discourse analysis and corpus linguistics approach." *Journal of Multilingual and Multicultural Development* 45(10): 4448-4467.
- Maamujav, U. (2021). "Examining lexical features and academic vocabulary use in adolescent L2 students' text-based analytical essays." *Assessing Writing* 49: 100540.
- Matthew, O. A., et al. (2024). "Discourse markers, coherence and its impact on cognitive interpretation of texts." *International Journal of Literature, Language and Linguistics* 7(2): 28-37.
- MISNAWATI, M., et al. (2024). "Corpus linguistics today: A qualitative approach." *Research and Innovation in Applied Linguistics* 2(1): 45-62.
- Pan, M. X. and Y. Zhu (2022). "Researching English language textbooks: A systematic review in the Chinese context (1964–2021)." *Asian-Pacific Journal of Second and Foreign Language Education* 7(1): 30.
- Pascual, D., et al. (2023). Digital scholarly practices in scientific communication: Paths and goals in research dissemination. *Digital scientific communication: Identity and visibility in research dissemination, Springer*: 3-30.

- Qazi, M. H., et al. (2023). "Representation of indigenous languages employing a religious screen for the discursive construction of students' postcolonial national identities: A curious case of 'internal colonisation' and 'cultural invasion' in Pakistani schools." *British educational research journal* 49(1): 35-52.
- Ren, X. (2022). "Evaluating Lexical Complexity From EGP to ESP Textbooks." *Journal of Language Teaching & Research* 13(5).
- Sampson, G. (2024). *Structural Linguistics in the 21st Century*, Cambridge Scholars Publishing.
- Saulais, P. (2023). *Knowledge and Ideation: Inventive Knowledge Analysis for Ideation Stimulation*, John Wiley & Sons.
- Stella, F. (2021). "Digital Philology and Quantitative Criticism of Medieval Literature: Unconventional Approaches to Medieval Latin Literature II."
- Uchida, S. and M. Negishi (2025). "Assigning CEFR-J levels to English learners' writing: An approach using lexical metrics and generative AI." *Research Methods in Applied Linguistics* 4(2): 100199.