

COMPARING THE LEXICAL LOAD IN THE ESL TEXTBOOKS FOR GRADE SEVEN: A CORPUS-BASED STUDY

Mahnoor Fatima

MPhil Scholar, Department of English, University of Sargodha, Punjab, Pakistan.

Email: mahnoorf809@gmail.com

Azhar Pervaiz

Associate Professor, Department of English Language and Linguistics, University of Sargodha, Pakistan. Email: azhar.pervaiz@uos.edu.pk

Abstract

This study investigates the lexical load of two Grade Seven English as a Second Language (ESL) textbooks widely used in Pakistan, one published by the Punjab Curriculum and Textbook Board (PCTB) for public schools and the other by Oxford University Press (OUP) for private institutions. The research aims to provide a comparative lexical profile of the two textbooks, focusing on frequency distribution of word types and tokens, the balance between content and function words, and the alignment of vocabulary with established frequency lists such as the General Service List (GSL), the Academic Word List (AWL), and the British National Corpus–Corpus of Contemporary American English (BNC–COCA) 25 frequency levels. The study employed corpus-based methods to analyze the vocabulary load, coverage, and lexical diversity across both textbooks. Findings reveal significant variation in lexical frequency distribution and the extent to which the vocabulary in each textbook supports learners' academic and communicative competence. The results highlight the implications for material selection, curriculum design, and vocabulary teaching in the Pakistani ESL context, pointing to the need for greater standardization and evidence-based textbook development.

Keywords: *ESL textbooks, lexical profiling, vocabulary load, corpus linguistics, Pakistan.*

Introduction

English has become one of the most important languages in the modern world. It is used as a medium of communication in education, administration, science, technology, and international trade. In Pakistan, the importance of English is even greater because it holds the status of an official language and is widely considered a symbol of prestige and opportunity (Rahman, 2004; Shamim, 2011). Knowledge of English opens doors to higher education, employment, and social mobility, making it one of the most desired skills among students and professionals. For this reason, English has been introduced as a compulsory subject in schools from the very beginning of formal education. However, despite these measures, the overall proficiency of Pakistani learners in English remains uneven. A large number of students, particularly those from public schools, still struggle with the language and are unable to communicate effectively.

This situation can be linked to many factors, including policies, teaching practices, and motivation. Some researchers argue that inconsistent policies have hindered progress in English education (Mansoor, 2005), while others highlight weaknesses in classroom teaching methods (Shamim, 2000) and learner motivation. These are important aspects, but there is another factor that often does not get enough attention: the textbook. In Pakistan, the textbook plays a decisive role because, for many students, it is the only source of English exposure. Unlike learners in countries where English is present in the environment, Pakistani students often encounter the language only inside the classroom. This makes the content of the textbook crucial for shaping their vocabulary, reading skills, and overall language ability.

In public schools, English is taught through textbooks published by the Punjab Curriculum and Textbook Board (PCTB). These books are provided free of cost to students and are designed under government supervision. For many learners, especially those from less privileged backgrounds, the PCTB textbook is the only exposure to English they receive. On the other hand, private schools often adopt textbooks published by Oxford University Press (OUP), which follow different approaches and are usually considered more advanced. This dual system creates a clear divide in the quality of learning opportunities, as students in private schools are believed to have better resources and exposure. Despite this, there has been little systematic research comparing the lexical content of these textbooks to see how they actually differ in terms of vocabulary input.

Vocabulary is one of the most important aspects of language learning. Without a sufficient vocabulary base, learners cannot understand texts, participate in discussions, or express themselves effectively. Researchers have shown that vocabulary knowledge is directly linked to language proficiency, and in many ways, learning vocabulary is equal to learning a language (Hazenberg & Hulstijn, 1996; Hu & Nation, 2000). To achieve comprehension, learners need to know at least 95 percent of the words in a text, and for unassisted understanding, they need to know about 98 percent (Laufer, 1997; Hu & Nation, 2000). This means that textbooks must be designed carefully to introduce the most frequent and useful words, recycle them across units, and gradually expand the learners' vocabulary to include academic words needed for advanced learning. If textbooks fail in this task, learners will not reach the lexical threshold required for comprehension, and their progress in English will be limited.

For this reason, researchers increasingly rely on corpus-based tools and frequency lists to evaluate the vocabulary load of textbooks. Standard lists such as Michael West's (1953) General Service List (GSL) of 2,000 high-frequency word families and Coxhead's (2000) Academic Word List (AWL) of 570 items are widely used benchmarks. More recently, the British National Corpus and Corpus of Contemporary American English (BNC/COCA) have been divided into 25 frequency bands of 1,000 words each. Studies show that learning the most frequent 3,000 words gives learners around 89 percent coverage of ordinary English texts (Nation & Waring, 2004; Schmitt & Schmitt, 2014). These frequency lists and bands provide reliable standards against which the vocabulary of textbooks can be measured.

In this study, the Grade Seven textbooks from PCTB and OUP were analyzed against these lists using corpus-based tools such as AntConc, AntWordProfiler, and Compleat Lexical Tutor. The analysis focused on four main aspects: the distribution of word tokens and types, the ratio of content words to function words, the alignment of vocabulary with the GSL and AWL, and the spread of words across the 25 BNC/COCA frequency bands. Together, these measures provide a clear picture of how much support the textbooks give to learners in building a strong vocabulary. The importance of such an analysis lies in the fact that textbooks in Pakistan are not often developed with corpus data in mind. Vocabulary selection is sometimes based on intuition rather than systematic criteria, which leads to uneven coverage and weak recycling of important words (Nation, 2001; Koprowski, 2005; Gouverneur, 2008). International studies show that such practices hinder vocabulary development because learners do not meet essential words often enough to retain them (Abello-Contesse & López-Jiménez, 2010). In Pakistan, the problem is even more serious because opportunities for incidental learning outside school are limited. This makes it even more necessary to evaluate the textbooks that learners depend on.

The objectives of this study were:

- a) To calculate the number of words introduced in both the PCTB and OUP Grade Seven textbooks.
- b) To analyze the vocabulary of these books against the 25 levels of BNC/COCA to see which frequency bands are represented.
- c) To examine the ratio of content and function words in each textbook.
- d) To compare the lexical profiles of the two textbooks to identify similarities and differences.

From these objectives, the following research questions were developed:

1. What is the lexical frequency distribution of word types and tokens in the OUP and PCTB Grade Seven ESL textbooks?
2. What is the ratio of content words to function words in the OUP and PCTB textbooks, and what does it indicate about their lexical load?
3. To what extent do the vocabulary items in the OUP and PCTB textbooks align with the General Service List (GSL) and the Academic Word List (AWL)?
4. How do the OUP and PCTB textbooks distribute vocabulary across the 25 frequency bands of the BNC/COCA corpus?

By answering these questions, the study provides a detailed picture of the vocabulary input available to learners in public and private schools in Pakistan. It highlights whether the textbooks meet the required coverage for comprehension and how their lexical designs differ. The results are expected to contribute to both academic research and practical improvements in textbook design, ensuring that learners in Pakistan have access to the vocabulary needed for effective language learning.

Literature Review

English language teaching has always been an important concern in Pakistan because of its impact on education, career opportunities, and social mobility. English holds a special position as it functions not only as a compulsory school subject but also as the official language of the state. It is widely used in higher education, professional life, and administration, and this makes its learning vital for success in many fields (Rahman, 2004; Shamim, 2011). Despite the high status of English, the outcomes of English teaching remain uneven across the country. Students in private schools often achieve better proficiency than those in public schools, and this difference is shaped by the learning resources available to them. One of the most important resources in both systems is the textbook, which plays a decisive role in shaping learners' exposure to the language.

Research into the causes of weak English learning in Pakistan has highlighted different factors. Some studies point to problems in language policy, which have resulted in inconsistent decisions about the role of English in education (Mansoor, 2005). Others stress the shortcomings of classroom teaching methods and the lack of communicative practices (Shamim, 2000). Motivation has also been identified as a challenge, with many learners showing limited interest in English because of social, cultural, or personal reasons (Shams, 2008; Shahbaz et al., 2011). While all these factors are relevant, another important area that has not been explored in enough depth is the nature

of textbooks. Since textbooks provide the primary input for learners, especially in public schools where no supplementary resources are available, their quality and lexical content must be studied carefully.

In public schools, textbooks are produced and distributed by the Punjab Curriculum and Textbook Board (PCTB). These books are provided free of cost to students and are often the only material used in classrooms. Because of this, they occupy a central role in the teaching and learning of English for a very large number of learners in Punjab, the most populous province in Pakistan. In private schools, however, English is usually taught through textbooks published by Oxford University Press (OUP). These books are commercially produced, follow different pedagogical orientations, and are generally seen as more advanced in design and content. This dual system creates a significant difference in the type and amount of input learners from public and private schools are exposed to. It also raises important questions about the extent to which these textbooks are able to support vocabulary development, which is essential for second language acquisition.

Vocabulary has been described by many researchers as the foundation of language learning. Hazenberg and Hulstijn (1996) argued that vocabulary knowledge is central to comprehension, and Hu and Nation (2000) pointed out that without knowing at least 95 percent of the words in a text, learners cannot follow its meaning. For independent reading and comprehension, the threshold rises to 98 percent (Laufer, 1997; Nation, 2001). This makes vocabulary one of the most important elements of second language acquisition. Learners may be taught grammar and structures, but without a strong vocabulary base, their ability to communicate remains limited. Textbooks therefore have the responsibility of providing sufficient exposure to high-frequency words, academic vocabulary, and systematic recycling that helps learners retain what they learn. Unfortunately, vocabulary selection in textbooks is often not based on systematic principles. Nation (2001) observed that vocabulary in many language teaching materials is not introduced with reference to standard word lists or corpus-based evidence. Instead, authors may choose words based on personal judgment, leading to a lack of consistency and poor recycling of essential items. Koprowski (2005) and Gouverneur (2008) noted similar issues, showing that vocabulary input in textbooks can be haphazard rather than planned. This has serious consequences in contexts like Pakistan where learners have limited opportunities for incidental learning outside the classroom. If important words are not repeated enough, learners cannot retain them, and their progress in English slows down.

Researchers such as Abello-Contesse and López-Jiménez (2010) also stress the importance of recycling in vocabulary learning. Words need to be encountered multiple times and in different contexts for learners to acquire them fully. A single exposure in a textbook is not enough. This makes the design of textbooks especially important: they should be structured in a way that provides repeated opportunities for learners to engage with the most useful vocabulary. Without this, learners are unable to build the strong lexical foundation required for comprehension, communication, and academic success.

The increasing use of corpora in applied linguistics has changed the way vocabulary is studied and has also influenced the evaluation of textbooks. Standard word lists such as Michael West's (1953) General Service List (GSL) of 2,000 high-frequency word families and Coxhead's (2000) Academic Word List (AWL) of 570 academic words are widely used to measure vocabulary input. Together, these lists provide a strong benchmark for determining whether textbooks expose learners to the words they need most. More recently, the British National Corpus and Corpus of

Contemporary American English (BNC/COCA) have been divided into 25 bands of 1,000 words each. Studies have shown that the most frequent 3,000 words cover nearly 89 percent of ordinary English texts (Nation & Waring, 2004; Schmitt & Schmitt, 2014). This makes the BNC/COCA framework a practical tool for analyzing the vocabulary in textbooks and for identifying whether learners are being exposed to high-frequency words in sufficient quantity.

Researchers recommend that textbooks should be informed by these frequency lists. Schmitt (2008) and Nordlund (2016) argued that using corpus-based evidence ensures that learners encounter the most useful words early and often. Burton (2012) also highlighted that textbooks which incorporate frequency data give learners purposeful and meaningful input. However, studies applying lexical frequency profiling to textbooks have shown that many materials fall short. Cobb and Horst (2011) and Webb and Nation (2017) found that textbooks sometimes underrepresent academic vocabulary and fail to recycle high-frequency words adequately. This problem is not only international but also relevant to Pakistan, where textbooks are often developed without reference to corpus data.

Research on Pakistani textbooks has been limited, and most studies have focused on areas other than vocabulary. Mahmood et al. (2014), for example, examined ideological and cultural aspects of textbooks but did not analyze lexical load. Azim and García (2020) investigated some issues of vocabulary coverage but did not carry out a systematic comparative analysis between public and private school textbooks. More recently, Khan (2024) studied vocabulary in secondary-level materials and found weak alignment with international frequency lists. These studies suggest that while vocabulary in Pakistani textbooks is an important issue, it has not been explored in depth, and comparative data between different systems of education are still missing.

This lack of research is surprising given the scale of English language teaching in Pakistan and the central role of textbooks in the process. Public schools rely heavily on PCTB textbooks, while private schools depend on OUP materials. Millions of students across the country are influenced by these resources, yet there is no clear evidence about how their vocabulary compares in terms of frequency, recycling, and coverage. Without such information, it is difficult for policymakers, curriculum developers, and teachers to make informed decisions about improving English teaching.

The present study responds to this gap by carrying out a corpus-based lexical analysis of the PCTB and OUP Grade Seven textbooks. It examines their vocabulary against standard lists such as the GSL, AWL, and the BNC/COCA frequency bands, and also looks at the distribution of content and function words. By doing so, the study provides evidence about the adequacy of these textbooks in supporting learners' vocabulary growth. It also highlights differences between public and private school materials, which has important implications for educational equity. Learners in different systems may not be receiving the same quality of input, and this can reinforce existing inequalities in access to English and its associated opportunities. This research gap is highly significant. Textbooks are the only consistent input for most learners in public schools and one of the main sources of English in private schools. If these materials fail to provide adequate vocabulary, learners cannot meet comprehension thresholds, which directly affects their language growth. The present study addresses this gap by comparing the lexical content of Grade Seven textbooks published by PCTB and OUP. Using corpus-based tools, it examines the frequency distribution, coverage of high-frequency and academic vocabulary, and alignment with standard

frequency lists. In doing so, it seeks to evaluate whether these textbooks are designed in ways that genuinely support learners' acquisition of English vocabulary.

Methodology

This study was designed to carry out a comparative lexical analysis of two English textbooks prescribed for Grade Seven learners in Punjab, Pakistan. One textbook was published by the Punjab Curriculum and Textbook Board (PCTB) and is taught in all public schools and some private institutions, while the other was published by Oxford University Press (OUP) and is used only in some private schools. Both textbooks were chosen because they represent two different educational systems, one locally produced following national curriculum standards and the other designed by an international publisher with global reach. Together, they form the core input of English learning for thousands of learners across Punjab. Since textbooks in Pakistan often serve as the sole linguistic resource available to students, their vocabulary content is critical for language acquisition. By analyzing and comparing these textbooks, the study aimed to determine whether learners learning from different ESL textbooks are receiving equitable exposure to vocabulary that is necessary for comprehension, communication, and further academic development.

The research adopted a quantitative design, guided by the principles of corpus linguistics and lexical frequency profiling. Quantitative methods are suitable because they allow for systematic measurement of vocabulary load, type-token distribution, and frequency coverage in an objective and replicable manner. A corpus-based approach provides a reliable way of examining language patterns, reducing researcher bias and ensuring that findings are grounded in evidence. The theoretical foundation for this analysis was drawn from Nation and Laufer's (1995) Lexical Frequency Profiling (LFP) framework, which classifies vocabulary according to frequency bands drawn from established lists such as the General Service List (GSL), the Academic Word List (AWL), and the British National Corpus/Corpus of Contemporary American English (BNC/COCA 25). Originally developed to evaluate learner writing, the LFP approach has since been extended to the analysis of teaching materials. It highlights the importance of high-frequency vocabulary for comprehension and fluency, as well as the need to assess how much of this vocabulary is provided in instructional texts.

The two selected textbooks were digitized in their entirety to prepare separate corpora for computational analysis. All units, reading passages, and exercises were typed and carefully checked to remove errors in formatting and transcription. This process ensured that the digital versions accurately represented the lexical content of the original books. Once digitized, the corpora were processed using a set of established software tools. AntConc was used to calculate the overall number of word tokens and word types in each textbook. Word tokens represent the total number of running words, while word types represent the number of unique words. Together, these measures provide a picture of the size and variety of vocabulary in each text and help assess lexical richness.

AntWordProfiler was employed to categorize words into content words and function words. Content words include nouns, verbs, adjectives, and adverbs, which carry the main semantic load of a text, while function words such as articles, prepositions, pronouns, and auxiliary verbs serve grammatical purposes. The ratio between content and function words is an important indicator of lexical complexity, as a higher proportion of content words suggests greater cognitive demand for learners. An online part-of-speech tagging tool was also used to refine the classification of content

words into their grammatical categories. This step allowed a more detailed understanding of the balance of word classes in each textbook.

The analysis also measured the alignment of textbook vocabulary with standardized frequency lists. AntWordProfiler was used to compare the texts against the General Service List (West, 1953) and the Academic Word List (Coxhead, 2000), both of which are widely recognized benchmarks in applied linguistics. This made it possible to calculate what percentage of each textbook's vocabulary comes from high-frequency general words and what percentage consists of academic vocabulary. Compleat Lexical Tutor, a web-based platform developed by Cobb (2002), was then used to analyze the texts against the BNC/COCA 25 frequency bands. This tool profiles vocabulary across 25 levels of 1,000 words each, beginning with the most frequent words in English. It allowed the study to determine how many words in the textbooks came from the first 3,000 most frequent bands, which are crucial for comprehension, and how many belonged to less frequent bands or were off-list words. Off-list words, which fall outside standardized lists, are particularly significant because they may place unnecessary burden on learners without contributing much to their communicative competence.

The independent variable in this study was the type of textbook, whether published by PCTB or OUP, while the dependent variables were lexical features measured through profiling. These included the total number of word tokens and types, the type-token ratio, the distribution of content and function words, and the coverage of vocabulary according to the GSL, AWL, and BNC/COCA 25 frequency bands. By systematically measuring these features, the study aimed to compare the lexical demands placed on learners by each textbook and to assess how far the materials align with recognized standards of vocabulary acquisition.

Data analysis followed a descriptive statistical approach. The results generated by AntConc, AntWordProfiler, and Compleat Lexical Tutor were tabulated for clarity. For example, the presence or absence of academic vocabulary, the balance between high- and low-frequency words, and the amount of vocabulary recycling could be easily observed through comparative tables. Ethical considerations were minimal because the study analyzed publicly available textbooks used in schools. Since the research was entirely based on secondary data in the form of textbooks, formal ethical approval was not required.

The methodology employed a quantitative, corpus-based approach to evaluate the lexical features of Grade Seven ESL textbooks published by PCTB and OUP. Through the use of AntConc, AntWordProfiler, Compleat Lexical Tutor, and POS tagging, the study systematically measured word tokens, word types, content-function ratios, and frequency coverage against established benchmarks such as the GSL, AWL, and BNC/COCA 25. The findings generated through this process provide a reliable and replicable basis for comparing the lexical load of the two textbooks and for drawing conclusions about the adequacy of vocabulary load offered to learners in Pakistan's public and private schools.

Analysis and Discussion

This study examines the lexical characteristics of two Grade Seven ESL textbooks: one published by Oxford University Press (OUP) and the other by the Punjab Curriculum and Textbook Board (PCTB). The analysis addresses four dimensions of lexical complexity: (1) frequency distribution of word tokens and types, (2) ratio of content words to function words, (3) alignment with the General Service List (GSL) and Academic Word List (AWL), and (4) distribution across the 25 frequency bands of the BNC-COCA corpus.

The first research question examined overall lexical input through word tokens (total words) and word types (unique words). The OUP textbook contains 29,853 tokens and 5,336 types, whereas PCTB contains 10,852 tokens and 2,770 types.

Table 1

Difference in word-token, word-types, and type-token ratio in PCTB and OUP textbooks

Metric	OUP Textbook	PCTB Textbook
Total Word Tokens	29,853	10,852
Total Word Types	5,336	2,770
Type-Token Ratio (TTR)	0.179	0.255

The nearly threefold difference in word count indicates that OUP provides learners with greater exposure to lexical items, which is a crucial factor for effective vocabulary acquisition. Repeated exposure to words enhances retention and internalization of vocabulary. Zahar et al. (2001) demonstrated that repeated encounters with the same lexical items significantly improve learners' retention. Milton (2009) emphasized frequent repetition as essential for L2 vocabulary development. Schmitt (2010) highlighted that the frequency of lexical encounters strongly influences acquisition and retention. The extensive lexical input in OUP textbook aligns with these findings, offering learners multiple opportunities to reinforce form-meaning connections.

From a theoretical perspective, Nation's Four Strands Model (2001) stresses the importance of meaning-focused input alongside deliberate vocabulary practice. The OUP textbook operationalizes this principle by providing extensive, contextually varied input, supporting incidental vocabulary learning and internalization. In contrast, PCTB offers a smaller corpus with limited repetition, which may constrain retention and mastery of new vocabulary. Krashen's Input Hypothesis (1985) further underscores that learners benefit from input slightly beyond their current level, suggesting that OUP's broader and repeated lexical exposure is pedagogically advantageous. The Type-Token Ratio (TTR) offers additional insight into lexical richness. PCTB textbook exhibits a TTR of 0.255, while OUP shows 0.179. At first glance, PCTB textbook appears more lexically diverse; however, TTR is inversely related to text length. Shorter texts often show artificially high TTR due to limited word repetition (Meara & Bell, 2001; Malvern & Richards, 2002). Therefore, OUP textbook's lower TTR reflects deliberate repetition—a beneficial feature for L2 vocabulary consolidation. Repeated encounters in varied contexts, as in OUP textbook, promote robust mental lexicon development. The PCTB textbook's higher TTR indicates breadth without sufficient reinforcement, potentially imposing higher cognitive demands.

Overall, the analysis highlights distinct pedagogical designs. OUP textbook prioritizes depth and reinforcement, supporting retention and mastery, whereas PCTB textbook emphasizes breadth within a limited word count. Laufer's (1997) Lexical Threshold Hypothesis suggests that learners require knowledge of approximately 3,000 word families for 95% comprehension and 5,000 families for 98% comprehension. With nearly 30,000 words and over 5,000 types, OUP textbook supports learners in reaching these lexical thresholds, facilitating independent comprehension and language development.

The second research question examines the ratio of content words to function words, a key indicator of lexical load. Content words (nouns, verbs, adjectives, adverbs) carry semantic meaning, whereas function words primarily serve grammatical purposes.

Table 2

Comparison of content-words and function-words in PCTB and OUP Textbooks

Category	OUP Tokens	OUP Types	PCTB Tokens	PCTB Types
Function Words	13,450	91	4,228	91
Content Words	16,403	5,245	6,624	2,679

The OUP textbook contains 16,403 content words and 13,450 function words (1.22:1 ratio), while PCTB contains 6,624 content words and 4,228 function words (1.56:1). Although PCTB shows a higher ratio, this does not necessarily indicate greater lexical complexity. Lexical load depends on text length, vocabulary variety, and repetition, not only on ratios. OUP's larger corpus offers more repeated encounters with content words, reinforcing retention and promoting deeper processing (Nation & Webb, 2011). This supports mental lexicon development, where repeated exposure consolidates vocabulary and facilitates automaticity.

Table 3

Category	OUP Textbook	PCTB Textbook
Noun	9,011	3,512
Verb	3,965	2,083
Adjective	2,524	755
Adverb	903	274

Comparison of content-words categories in PCTB and OUP textbooks for Grade Seven

Analysis of content word distribution indicates that OUP textbook contains 9,011 nouns, 3,965 verbs, 2,524 adjectives, and 903 adverbs, whereas PCTB textbook contains 3,512 nouns, 2,083 verbs, 755 adjectives, and 274 adverbs. Nouns dominate both textbooks, reflecting the descriptive and expository nature of educational texts (Biber et al., 1998). The higher verb count in OUP emphasizes action and sentence construction skills (Nation, 2001), while greater presence of adjectives and adverbs suggests focus on descriptive and elaborative abilities. Ellis (1994) argued that repeated exposure to content words strengthens learners' mental lexicon, enhancing expressive and academic language development.

In contrast, PCTB textbook provides a limited, foundational lexical set. While this supports essential vocabulary acquisition, restricted variety may impede expressive language development and academic skill growth. The analysis indicates that OUP promotes a more balanced lexical environment, combining breadth, repetition, and variety across grammatical categories.

The third research question explores alignment with frequency-based lists, reflecting the textbooks' support for general and academic vocabulary.

Table 4

Comparison of GSL coverage in OUP and PCTB Textbooks for Grade Seven

Category	OUP Textbook	PCTB Textbook
Total Word Tokens	29,853	10,852
GSL Tokens	24,013	8,523
GSL Coverage %	80.44%	78.54%

Table 4 shows that OUP textbook contains 24,013 GSL tokens (80.4% of total words), while PCTB textbook contains 8,523 GSL tokens (78.5%). Both textbooks provide substantial coverage of high-frequency words, aligning with Nation (2006), who recommends at least 80% GSL coverage for adequate comprehension. OUP textbook's slightly higher coverage and repeated exposure support greater retention and foundational proficiency.

Table 5

Comparison of AWL coverage in PCTB and OUP textbooks for Grade Seven

Metric	OUP Textbook	PCTB Textbook
Total Content Word Tokens	16,403	6,624
Total Content Word Types	5,245	2,679
AWL Word Tokens Found	865	315
AWL Word Types Found	187	134
AWL Token Coverage (%)	5.27%	4.75%
AWL Type Coverage (%)	3.57%	5.01%

Regarding academic vocabulary, OUP textbook contains 865 AWL tokens and 187 unique types (5.27% token coverage, 3.57% type coverage), while PCTB textbook contains 315 tokens and 134 types (4.75% token coverage, 5.01% type coverage). This suggests that OUP textbook prioritizes repeated reinforcement of AWL items, enhancing retention and facilitating academic readiness. The PCTB textbook presents a broader but less reinforced academic set, supporting recognition but limiting long-term retention.

Early exposure to academic vocabulary in middle school supports learners' transition to secondary education (Coxhead, 2000; Hyland & Tse, 2007). Webb (2007) highlights repetition frequency as a critical factor for retention, indicating that OUP textbook's method of integrating AWL tokens systematically is more pedagogically effective. Both textbooks address core vocabulary needs, but OUP textbook demonstrates superior depth and reinforcement, supporting sustained lexical growth.

The fourth research question examines the distribution of vocabulary across the 25 frequency bands of the BNC-COCA corpus, providing insight into coverage of high-, mid-, and low-frequency words.

Table 6

Comparison of overall coverage of BNC-COCA 25 in PCTB and OUP textbooks for Grade Seven

Metric	Oxford Textbook	PTB Textbook
Total Word Tokens	29,853	10,852
Total Coverage (%)	98.91%	98.24%

Table 6 shows OUP textbook achieves 98.9% coverage, while PCTB textbook achieves 98.2%. Both surpass the 98% optimal threshold suggested by Laufer & Ravenhorst-Kalovski (2010) and Nation (2001), ensuring sufficient comprehension for ESL learners.

Table 7

Comparison of vocabulary coverage in top5 bands of BNC-COCA25 in PCTB and OUP textbooks for Grade Seven

Frequency Band	OUP Tokens	OUP %	OUP cumul %	PCTB Tokens	PTB %	PCTB Cumul %
K1	25,017	83.80%	83.80%	8,710	80.26%	80.26%
K2	2,142	7.17%	90.97%	1,005	9.26%	89.52%
K3	837	2.80%	93.77%	480	4.42%	93.94%
K4	528	1.76%	95.53%	139	1.28%	95.22%
K5	245	0.82%	96.36%	108	1.00%	96.22%

Table 7 shows that most vocabulary in both textbooks comes from the top five bands (K1–K5), with cumulative coverage of 96.36% in OUP textbook and 96.26% in PCTB textbook. This focus on high-frequency words aligns with Zipf's Law (1949) and research by Biber & Reppen (2015), emphasizing that natural language is dominated by a small set of highly frequent words. The predominance of these words ensures learners acquire core vocabulary essential for comprehension, fluency, and academic readiness (Adolphs & Schmitt, 2003).

Beyond the top five bands, OUP textbook includes vocabulary from mid- and low-frequency bands (K6–K25), supporting gradual lexical expansion and literacy development (Yildiz, 2023). PCTB textbook in contrast, emphasizes accessibility and high-frequency words, limiting exposure to advanced lexical items. This distinction highlights a fundamental pedagogical difference. The OUP textbook balances reinforcement of core vocabulary with gradual introduction of less frequent words, while PCTB textbook prioritizes foundational comprehension.

The comparative analysis reveals that the OUP textbook provides broader lexical input, repeated exposure, and richer variety across word classes. It supports vocabulary retention, academic skill development, and cognitive engagement. Its alignment with GSL and AWL, and inclusion of mid- and low-frequency vocabulary, equips learners for independent reading and content-based learning.

PCTB textbook ensures accessibility and presents a smaller corpus with limited reinforcement. It offers foundational vocabulary without extensive repetition. Consequently, retention, expressive language development, and academic growth may be constrained. These findings highlight the importance of input quantity, repetition, and strategic lexical selection in ESL materials, as emphasized by Nation (2001, 2013), Krashen (1985), and Laufer (1997).

The OUP textbook operationalizes these theoretical principles by providing repeated encounters with high-frequency vocabulary, integrating academic words systematically, and gradually expanding lexical range beyond high-frequency bands. PCTB textbook primarily supports foundational vocabulary recognition without promoting extensive lexical development. From a pedagogical standpoint, OUP textbook fosters both depth and breadth. The PCTB textbook emphasizes breadth with limited depth, reflecting differing priorities in textbook design.

The findings underscore the role of textbook design in shaping lexical acquisition. Repetition and exposure to diverse word classes, as seen in OUP textbook facilitate robust mental lexicon formation, vocabulary retention, and expressive language development (Schmitt, 2008; Ellis, 1994). Exposure to high-frequency vocabulary ensures comprehension and functional language use, while mid- and low-frequency words provide opportunities for lexical expansion and academic growth (Yildiz, 2023; Adolphs & Schmitt, 2003). OUP textbook's approach aligns with best practices in vocabulary pedagogy, combining frequency, repetition, variety, and progressive complexity, whereas PCTB offers a more limited, foundational approach.

Conclusion

The findings provide insights into the lexical load, breadth, and pedagogical design of the two instructional materials and highlight their implications for curriculum, pedagogy, and textbook development.

The analysis revealed significant differences in lexical input between the textbooks. The OUP textbook contains 29,853 word tokens and 5,336 word types, compared to 10,852 tokens and 2,770 types in the PCTB textbook [insert Table 4.1]. This indicates that learners using OUP are exposed to a broader and more varied vocabulary, while PCTB focuses on a compact, foundational lexical set. Although PCTB demonstrates a higher type-token ratio, this is largely due to its smaller word count and limited repetition, which may limit long-term retention. In contrast, the OUP textbook facilitates vocabulary acquisition through repeated exposure, promoting retention, inferencing skills, and lexical flexibility (Grabe & Stoller, 2019; Van Zeeland & Schmitt, 2013).

Analysis of content and function words further highlights differences in lexical load. The OUP textbook contains more content words across all grammatical categories, including nouns, verbs, adjectives, and adverbs, whereas PCTB relies on simpler structures with fewer opportunities for semantic expansion. This suggests that OUP provides richer input for developing expressive and academic language skills. Both textbooks achieve high coverage of high-frequency vocabulary (K1–K5 bands), meeting the 95–98% comprehension benchmark. However, OUP extends coverage into mid- and low-frequency bands, supporting incremental lexical growth and advanced language proficiency, while PCTB emphasizes foundational vocabulary for accessibility and minimal cognitive burden.

In terms of alignment with standard vocabulary lists, both textbooks provide substantial GSL coverage (~80%), ensuring learners acquire essential words. OUP additionally reinforces AWL items through repeated exposure, supporting academic literacy and vocabulary retention. PCTB introduces a wider variety of AWL items but with less repetition, promoting recognition but

limiting reinforcement. These findings suggest that effective ESL textbooks should balance breadth, repetition, and cognitive scaffolding to optimize vocabulary acquisition and comprehension.

The study has several practical implications. Curriculum and textbook developers should design materials that balance accessibility with lexical depth, strategically recycle core and academic vocabulary, and progressively introduce mid- and low-frequency words. Teachers should be trained to understand lexical profiles, supplement classroom instruction to reinforce vocabulary, and address gaps in textbook content (Azim & García, 2020; Richards, 2001). Future research should expand the scope to other grades, regions, and publishers, and include longitudinal and qualitative analyses of vocabulary, collocations, and contextual richness to measure the impact of lexical input on learners' language development.

In conclusion, this study highlights the importance of corpus-informed, frequency-based, and scaffolded vocabulary selection in ESL textbook design. The OUP textbook demonstrates an optimal combination of repetition, lexical breadth, and academic support, whereas PCTB provides foundational accessibility suitable for basic language learning. Aligning curriculum, pedagogy, and research-based lexical profiling can enhance learners' proficiency and academic success, offering a model for effective, evidence-based ESL instructional material development in the Pakistani context.

References

- Abello-Contesse, C., & López-Jiménez, M. D. (2010). Recycling vocabulary in second language learning: A study of L2 learners of English in Andalusia, Spain. *International Journal of English Studies*, 10(1), 41–62. <https://doi.org/10.6018/ijes/2010/1/113191>
- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425–438. <https://doi.org/10.1093/applin/24.4.425>
- Anthony, L., & Nation, I. S. P. (2017). *AntWordProfiler (Version 1.4.1) [Computer software]*. Waseda University.
- Azim, M. U., & Maldonado García, M. I. (2020). Lexical load of Punjab Curriculum and Textbook Board's English 1 and English 2. *Bulletin of Education and Research*, 42(2), 15–32.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1998). *Longman grammar of spoken and written English*. Longman.
- Cobb, T. (2002). *The compleat lexical tutor* [Website]. <https://www.lex tutor.ca>
- Cobb, T., & Horst, M. (2011). Does word coach coach words? *Applied Linguistics*, 32(2), 232–246. <https://doi.org/10.1093/applin/amq021>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213–238. <https://doi.org/10.2307/3587951>
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press.
- Gouverneur, C. (2008). The phraseological patterns of MAKE and DO: A corpus-based analysis. *International Journal of Corpus Linguistics*, 13(3), 223–258. <https://doi.org/10.1075/ijcl.13.3.02gou>
- Hazenbergh, S., & Hulstijn, J. H. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics*, 17(2), 145–163. <https://doi.org/10.1093/applin/17.2.145>
- Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430.
- Khan, A. F. (2024). Lexical Load in Class X English Textbooks: A Corpus-Based Comparative Analysis. *Linguistics and Literature Review*, 10(1), 122–139. <https://doi.org/10.32350/llr.101.07>
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

- Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp. 20–34). Cambridge University Press.
- Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, 19(1), 85–104. <https://doi.org/10.1191/0265532202lt221oa>
- Mansoor, S. (2005). *Language planning in higher education: A case study of Pakistan*. Oxford University Press.
- Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, 16(3), 5–19.
- Meara, P., & Suárez García, M. (2010). V_Yes!: A test of productive vocabulary. *Reading in a Foreign Language*, 22(1), 236–246.
- Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Nation, I. S. P., & Laufer, B. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Nation, I. S. P., & Waring, R. (2004). Vocabulary size, text coverage and word lists. In N. Schmitt (Ed.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 6–19). Cambridge University Press.
- Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary*. Boston: Heinle, Cengage Learning.
- Nordlund, M. (2016). EFL textbooks for young learners: A comparative analysis of vocabulary. *Education Inquiry*, 7(1), 47–68. <https://doi.org/10.3402/edui.v7.27564>
- Rahman, T. (2004). *Language policy and localization in Pakistan: Proposal for a paradigm shift*. Social Policy and Development Centre.
- Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge University Press.
- Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. <https://doi.org/10.1177/1362168808089921>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47(4), 484–503. <https://doi.org/10.1017/S0261444812000018>
- Shamim, F. (2000). Functional English at the intermediate level in Pakistan: A mismatch between policy and practice. *Language, Culture and Curriculum*, 13(2), 168–181. <https://doi.org/10.1080/07908310008666597>
- Shamim, F. (2011). English as the language for development in Pakistan: Issues, challenges and possible solutions. *British Council*.
- Tomlinson, B. (2013). *Developing materials for language teaching* (2nd ed.). Bloomsbury Academic.
- Van Zeeland, H., & Schmitt, N. (2013). Incidental vocabulary acquisition through L2 listening. *Studies in Second Language Acquisition*, 35(3), 483–517.
- Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford University Press.
- West, M. (1953). *A general service list of English words*. Longman.
- Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, 57(4), 541–572. <https://doi.org/10.3138/cmlr.57.4.541>