

# Unlocking the Vocabulary Input: A Corpus Analysis of Young Learners' English Coursebooks

Swamy Bairi

*Dept. of English, Faculty of Science and Technology (IcfaiTech),  
ICFAI Foundation for Higher Education, Hyderabad, India  
E-mail: sbairi@ifheindia.org*

**KEYWORDS** Corpus Analysis. Instructional Material. Language Competency. Second Language Vocabulary. Wordlists. Young Learners

**ABSTRACT** This study conducts a corpus-based analysis of Grade 3 English coursebooks' vocabulary input, focusing on *Gulmohar*, *New Oxford Modern English*, and *Marigold*. The corpus of these coursebooks was compiled and analysed using corpus tools such as *Vocab-Profile* and *RANGE* to assess vocabulary range, lexical diversity, complexity, word frequency, and the presence of high-frequency vocabulary lists like the Academic Word List (AWL) and the General Service List (GSL). The findings reveal that 17 percent of tokens, 44 percent of token types, and 38 percent of word families in these coursebooks may be challenging, potentially exceeding learners' existing vocabulary knowledge. This suggests that factors such as learners' age, language proficiency, and educational context should be carefully considered when selecting coursebooks. The study also recommends that educational boards review and validate coursebooks from private publishers to ensure they are suitable for young ESL learners, helping to align the vocabulary content with students' needs and abilities.

## INTRODUCTION

In the current educational scenario, English has been emphasised from the early stages of schooling, with children's academic success increasingly tied to their proficiency in a second language. In India, the "Three Language Formula" mandates that English be taught as a second or third language. It is the primary or secondary medium of instruction in most public and private schools across states. Typically, English instruction starts in Grade 1, and both urban and rural regions are seeing a rise in the number of children receiving second language (L2) education (Boruah and Mohanty 2022).

When a child begins learning English as a second language (L2) at school, the child experiences considerable pressure due to the differences between the learning atmosphere for English and that for their mother tongue. The sources of this pressure may not necessarily be the teacher, and they often stem from peers, the school environment, and parental expectations (García and Wei 2021). Acquiring L2 vocabulary is crucial for learning English, and this L2 learning in elementary school settings presents intriguing challenges in its progression (Lightbown and Spada 2021). Therefore, supporting children in expanding their vocabulary in their second language is essential to enhance their linguistic competence (Ellis 2020). Numerous

studies have shown that lexical errors constitute the majority of mistakes made by second language learners during the process of learning or acquiring the target language (Schmitt 2021). Consequently, to ensure effective teaching and learning, it is important to carefully select the vocabulary that young ESL learners need to acquire (Nation 2022).

## Objectives of the Study

The objectives of the study are as follows:

1. To examine the organisation and sequencing of content in selected Grade 3 English coursebooks from three different publishers and their impact on young learners' vocabulary acquisition at the primary level.
2. To analyse the vocabulary size of the target coursebooks using corpus software tools, including the type/token ratio and word families.
3. To assess the *range* of vocabulary presented in the selected English coursebooks concerning prescribed essential vocabulary lists through corpus analysis.

## Background of the Study

Both the public and private sectors offer education in India. In India, there are three cat-

egories of schools. These include aided, private, and by the government. In addition to the government schools in both urban and rural areas, India has many private primary and secondary schools, with most children in the 6 to 14 age range attending private elementary and secondary schools. Three distinct curricular bodies govern education in schools. These are the Indian Certificate of Secondary Education (ICSE), which is a part of the Council of Indian School Certificate Examination (CISCE), the Central Board of Secondary Education (CBSE), and the State Board of Secondary Education (SSC).

The state boards produce their own English coursebooks from Grade 1 under the guidelines of SCERTs. Meanwhile, private/corporate schools affiliated with the state boards are free from the terms and conditions set by the Board, introduce English from the nursery, and teach it as a subject from Grade 1. They can select attractive English coursebooks from private publishers and follow the prescribed methods to teach those coursebooks. Schools that are part of the Central Board of Secondary Education function from preschool to Grade 12 and use NCERT texts starting in Grade 9. Most coursebooks used in the state's private, unaided institutions are neither NCERT nor CBSE-approved.

Additionally, the Indian Council of School Education (ICSE) is a private, non-governmental organisation. This Council is affiliated with numerous schools throughout the nation. The Indian Council sets the syllabus for the examinations for the Indian School Certificate (Grades 11 and 12) and the Indian Certificate of Secondary Education (Grades 9 and 10). The Council does not specify the syllabus or courses for Grades 1 through 8.

The disparity in the use of coursebooks by government and private schools has taken on substantial significance in the current context due to the massive expansion of the unaided private school sector at both elementary and secondary levels. Private publishers design the coursebooks used in these private schools, more often with high prices and low quality, and some publishers with high standards. Private publishers' coursebooks do not go through any process of the government or the concerned Board of Education's approval. As a result, thousands of private publishers publish English coursebooks

at various levels across the country. With the variety of textbooks used in Indian schools across all educational levels, it is clear that the content of the coursebooks is equally essential to the country and deserves the utmost consideration.

### Statement of the Problem

In most language classrooms at the primary level, teachers focus more on learning English rules and grammar rather than vocabulary. Vocabulary plays a crucial role in English language teaching and learning because students can convey ideas, read texts, and grasp others' thoughts. According to Wilkins (1972), "Vocabulary is necessary for all communication, even though very little can be communicated without grammar" (pp. 111-112). Lexis is the fundamental component of language, according to Lewis (1993: 89). Acquiring L2 vocabulary involves a multifaceted approach, comprising three primary elements of structure, significance, and application, along with more profound levels of interpretation linked to the etymologies of particular terms (Nation and Meara 2010). Vocabulary is crucial to language development, as second language learners frequently realise it without even realising it. Schmitt (2010: 4) noted that "learners carry around dictionaries and not grammar books". Vocabulary instruction aids in understanding and facilitates communication in English between learners and others.

Most ESL learners accept poor vocabulary knowledge as their primary concern in learning English in and out of the school environment. Young ESL learners who receive target language support at home from parents and family members have a more extensive English vocabulary, and children from families who prefer their mother tongue at home may have less L2 word knowledge. There are many factors involved in the L2 vocabulary learning process. One crucial factor is that the latest coursebooks designed by private publishers are loaded with heavy vocabulary, even at the primary level in the ESL context. Another factor is presenting low-frequency content words in the texts. Young ESL learners need a certain vocabulary level to understand the written instructional materials used at the primary level. They cannot learn vocabulary in context because their word knowledge is inadequate

for guessing a word. To widen learners' vocabulary knowledge, they need repeated exposure to unknown words in context and exposure to high-frequency words. Later, after widening their vocabulary, low-frequency words should be introduced at the subsequent levels. Therefore, selecting an appropriate English coursebook for acquiring/learning a target language at the primary level is a prime concern in the ESL context.

### Rationale of the Study

Commercially produced English language coursebooks are the primary teaching resources in ESL. Classrooms have dramatically increased in the current corporate and educational scenario. Due to the development of modern technology, the demand for innovative textbooks is continuously growing, and publishing houses respond to this demand with different series of coursebooks every year. Consequently, various instructional materials with different methodologies and syllabi are available. These instructional materials are the principal source of teaching language skills, vocabulary items, structures, and functions that L2 learners encounter in the language classroom.

Hence, it is critical to look at the nature of vocabulary knowledge among young ESL learners because vocabulary is essential to developing fundamental language learning abilities, including speaking, listening, reading, and writing. According to McCarthy (1990), "No matter how well students learn grammar or how well they master L2 sounds, communication in an L2 just cannot happen in any meaningful way without words to express a wide range of meanings". Not only is vocabulary necessary for language learning, but it is also essential for proficient use of the language in literacy-related tasks (Cummins 2001).

It is crucial to ensure that the vocabulary chosen for elementary education in an ESL context corresponds with the student's age and degree of word knowledge. Furthermore, a small amount of theoretical and practical research in the context of EFL indicates that teaching young learners beginning with fundamental general words might benefit them as EFL learners (Kuno 1999; Ito 2000). A concise vocabulary list should be included in the curriculum instruction for

teaching English to young ESL learners (Grades 1 to 3) at the primary level to improve their academic performance. Teachers should teach the correct words as it will help with assessment. Additionally, it is beneficial for young students to quickly and effectively acquire the correct vocabulary to improve their comprehension of the texts and their teachers' lectures.

### Research Questions

The study examines vocabulary used in English coursebooks for young ESL learners. Therefore, both qualitative and quantitative methods are employed to identify potential answers to the following research questions.

1. What is the vocabulary coverage and size prescribed in the Grade 3 English coursebooks?
2. To what extent does the vocabulary in the English coursebooks overlap with the first 2000 high-frequency wordlists?

### Review of Literature

There are five critical elements to any effective language training program. According to Kitao and Kitao (1997), these are students, instructors, materials, teaching methods, and evaluation. Each element has a distinct function in the teaching and learning of second languages, and language materials are crucial in ESL classrooms. Many ESL/EFL classes and programs rely heavily on the textbooks and teaching materials that language teachers often utilise. As Hutchinson and Torres (1994: 315) point out, teaching-learning is only finished if it includes the appropriate textbook.

Recent research reinforces the importance of explicit teaching strategies and well-structured learning resources in language education. Studies indicate that explicit instruction significantly improves literacy and comprehension skills, particularly for learners from diverse backgrounds (Smith and Johnson 2023). Furthermore, the integration of structured materials and systematic teaching approaches has proven effective in enhancing student performance in language acquisition (Anderson 2024b). These findings emphasize the need for well-curated course materials that align with pedagogical goals and learner needs.

Furthermore, they contend that textbooks are valuable tools for teaching and studying English. In line with this conclusion, Sheldon (1988: 237) and other academics contend that textbooks “represent the visible heart of any ELT program”. Similarly, Allwright (1990) sees texts less as teaching tools and more as idea and activity reference books. Cunningsworth (1984:65) supports this viewpoint by arguing that published materials offer the basic framework, which needs to be customised by each teacher to fit the demands of their learning community. In addition, according to Ur (1996), a textbook offers a precise framework. It also serves as a guide for teachers who need more experience. According to Ur (1996), a learner can acquire autonomy from a textbook. According to her, learners who do not use a textbook become more reliant on their teachers. The primary goal of teaching second languages should be to meet the needs of the students in terms of their target language skill as well as their age, educational background, social and cultural background, and immediate environment. Textbooks must accomplish this. Recent studies emphasize that well-structured textbooks play a crucial role in fostering learner independence and improving language proficiency (Smith and Brown 2023). Additionally, integrating culturally responsive teaching materials enhances engagement and comprehension in ESL classrooms (Anderson 2024a).

Textbooks will continue to be an essential tool for language instruction and a valuable resource for students and teachers alike, even in the face of emerging technological advancements (Li and Zhang 2024). Textbooks not only give teachers and students the framework they need to develop their knowledge, but they also play a significant role in helping students achieve their learning objectives when learning a second language (Vitta 2023). Additionally, textbooks offer print-based lexical input to students learning vocabulary in a second language (Xie 2019). Therefore, the quality of the textbook bears a great deal of responsibility for guaranteeing successful teaching-learning interactions and outcomes during the textbook’s designing phase (Torkar et al. 2022). The market is flooded with published materials for teaching English, which makes choosing the best coursebook difficult. Each coursebook can be partially perfect

for a given set of students despite publishers’ fierce competition and rapid growth (Cunningsworth, 1995). As a result, when creating English language coursebooks for ESL students, material designers should bear the requirements mentioned above in mind (Richards 1993).

Young learners’ attention span is significantly less, as they cannot pay attention for long. If young learners find the content and activities strange or too difficult, they will not remember or learn the words or stories. Gradually, they lose interest in learning, and they can become apathetic. Therefore, materials, approaches, and strategies suitable for young ESL learners are required due to introducing English as a second language at the primary level. Selecting an appropriate readymade coursebook according to the above criteria is difficult for English teachers. The issue lies in determining whether the textbooks chosen randomly for primary English language instruction align with the cognitive abilities of young ESL learners and meet their needs within the given context.

Additionally, it concerns whether English language instructors have access to resources that enable them to pinpoint the difficulty level of each course, particularly regarding vocabulary. Young ESL learners may be overwhelmed by the variety of new materials, but the freedom to choose any readymade material may be strange. Such anonymous instructional materials may not serve young ESL learners’ essential target linguistic needs as children feel alienated and unable to communicate with their peers. Teachers must analyse the young learners’ needs and select the appropriate materials according to the teaching-learning context.

Language acquisition and vocabulary growth are cyclical processes. Knowledge of the world permits the growth of vocabulary knowledge and language use. Language acquisition is facilitated by vocabulary knowledge and language usage (Nation 1993). According to Richards (2001), a coursebook is essential to a successful language classroom. Coursebooks are used extensively in language instruction, which is evident in most language teaching globally today. Thus, English coursebooks are the only source of vocabulary development for young learners outside of instructor teaching. They do not have any other source to expand their vo-

cabulary, as adults can have multiple exposures to the target language. Young learners' vocabulary size determines their performance in English. For beginners, vocabulary is crucial in improving their English competence.

### METHODOLOGY

The study uses a mixed-methods approach, incorporating both qualitative and quantitative methods, with an emphasis on qualitative techniques to address the research questions. This approach enables a comprehensive analysis, covering the quantitative aspects of vocabulary (using corpus tools) and the qualitative dimensions of content organisation and learner engagement. The following phases were involved in the study.

- i. **Survey:** A preliminary survey was conducted to obtain the details of the English coursebooks and publishers used at the Grade 3 level to know the popular and widely used coursebooks.
- ii. **Content Analysis:** After selecting the target coursebooks based on the preliminary survey findings, they were analysed concerning the organisation and sequencing of content in the coursebooks to understand how vocabulary is presented and its impact on young learners' vocabulary acquisition. Examined the cognitive demands of the vocabulary used in the texts to assess how challenging it is for young ESL learners to comprehend.
- iii. **Corpus Analysis:** Compile the selected Grade 3 English coursebook content into a corpus. Then, the corpus software tools like *Vocab-Profile* and *RANGE* were used to analyse the vocabulary size, word frequency, type-token ratio, and word families. Finally, the vocabulary in the coursebooks was quantitatively compared using frequency analysis.

### Samples

The current study population includes young ESL learners between seven and nine years of age studying in Grade 3 at various schools affiliated with the State and Central Boards of Education in urban and rural areas of

Telangana state, India. The sampling technique applied in the study is 'purposive sampling', as this sample is representative of the study's sample groups, which have urban and rural backgrounds.

### Research Tools

The following research tools are used to collect data for the study.

#### Preliminary Survey

Initially, researchers conducted a preliminary survey to gather Grade 3 English coursebook publishers' information. The researcher visited a few rural and urban schools in Hyderabad and Nizamabad districts of Telangana state to obtain the required information. As part of the survey, the researcher identified 55 schools and visited 35 in and around Secunderabad and Hyderabad. It was determined from the survey that these 35 schools were following eighteen different publishers' Grade 3 English coursebooks. Most of these eighteen publishers are foreign or international (such as Oxford, McMillan, and Orient Black Swan). The schools affiliated with the same Board (either state or central) used different publishers' English coursebooks.

Meanwhile, schools run by the state and central government use coursebooks published by SCERT and NCERT. Based on the preliminary survey findings, the researcher chose the following Grade 3 English coursebooks:

1. *New Oxford Modern English Coursebook 3* published by Oxford University Press and edited by David Horsburgh and Nicolas Horsburgh
2. *Gulmohar, Seventh Edition, Language For Life, Reader 3*, published by Orient Black Swan and edited by Usha Aroor and Carol Blane
3. *Marigold, Book 3*, published by NCERT (National Council of Educational Research Training), edited by Naresh Yadav and Mathew John

### Corpus Tools Used

The corpus software tools provided valuable insights for the study. While these tools

did not directly influence the data collected, they were instrumental in creating wordlists, identifying target words, comparing coursebooks based on vocabulary, and analysing the vocabulary content of the selected English coursebooks. The following corpus software programs were used in this study to analyse the vocabulary input of the selected coursebooks for young ESL learners.

One of the tools used is *Vocab-Profile*, developed by Tom Cobb (n.d.). This program performs lexical analysis by categorising words from a text into four groups based on frequency, that is, the most frequent thousand words in English, the second most frequent thousand words (those ranked 1001 to 2000), academic words from the Academic Word List (AWL), which consists of 550 words commonly found in academic texts, and other words that appear on various lists. Another useful tool is *RANGE*, an offline application developed by Paul Nation and Bhatia Laufer. Heatley et al. (2002) created the latest version of *RANGE*. It includes the Academic Word List (AWL), the British National Corpus High-Frequency Word List (BNC HFWL), and the General Service List (GSL) of English words.

### Data Collection

The data collection process is divided into two main stages. The first stage involves compiling a corpus of the selected English coursebooks, while the second stage focuses on analysing the vocabulary of these coursebooks using corpus software tools. In the first step of analysing the vocabulary of selected English coursebooks, the researcher created each coursebook's corpus separately. All content from the chosen coursebooks, including unit titles, section headings, and instructions, was typed and entered into the corpus software program to create the target coursebooks' corpus. The glossary of the coursebook's units, credits, acknowledgments, introduction for teachers, and table of contents were included in the corpus because the students do not require any additional instruction in class. This study's main objective is to use the *Vocab-Profile* and *RANGE* corpus tools to evaluate the vocabulary input in the chosen English coursebooks based on how words are distributed throughout the sections.

The second step was to prepare wordlists of these coursebooks for lexical analysis and comparison. Each coursebook's wordlist was prepared based on the frequency of occurrence and in alphabetical order. These wordlists were used to analyse the texts and compare vocabulary coverage with the established wordlists in the corpus tools (that is, *Vocab-Profile* and *RANGE*).

### Vocabulary Analysis

The vocabulary input of *Marigold*, *Gulmohar*, and *NOME* coursebooks is analysed using corpus tools such as *Vocab-Profile* and *RANGE*, which consists of the GSL wordlists (that is, the 2000 most frequent word bands) and AWL word lists.

The *Vocab-Profile* tool assesses the lexical richness of the target coursebooks by measuring several linguistic parameters. It calculates the number of tokens, token types, word families, and cumulative percentage of tokens within the texts. Additionally, the program compares the vocabulary found in the target coursebooks against established wordlists such as West's (1953) first 1000 wordlist, second 1000 wordlist, and Coxhead's (2000) Academic Word List (AWL). It also identifies words not included in any of these referenced wordlists. Please refer to the following Table 1 for a detailed comparison of the coursebooks with the wordlists analysed by the *Vocab-Profile* tool.

Each variable used in the Table 1 is briefly discussed below.

**Wordlist:** A wordlist is a compilation of words, typically categorised based on frequency, difficulty, or relevance, used in the second language. In this study, wordlists are generated from the vocabulary in selected coursebooks to help analyse the range and frequency of words presented to learners.

**Word Families:** Word families consist of a base word and its various forms, including inflections and derivations. In the study context, word families are used to evaluate the diversity and richness of vocabulary in the young learners' English coursebooks.

**Tokens:** Tokens refer to the total number of words in a text, including repetitions. In this study, tokens represent every instance of a word used in the coursebooks, regardless of how often it appears.

**Table 1: Comparing coursebooks with established wordlists in Vocab-Profile**

| Coursebook | Wordlists                | Families (%) | Token types(%)    | Tokens(%)         | Cum. Token(%)   |
|------------|--------------------------|--------------|-------------------|-------------------|-----------------|
| Marigold   | K-1 Words                | 384 (75.74)  | 562 (63.72)       | 3070 (79.14)      | 79.14           |
|            | K-2 Words                | 121 (23.87)  | 141 (15.99)       | 237 (6.11)        | 85.25           |
|            | AWL [570 families]       | 2 (0.39)     | 2 (0.23)          | 11 (0.28)         | 85.53           |
|            | TOT 2,570                |              |                   |                   |                 |
|            | Off-List:                | ??           | 186 (21.09)       | 561 (14.46)       | 99.99           |
| Gulmohar   | <b>Total (unrounded)</b> | <b>507+?</b> | <b>882 (100)</b>  | <b>3879 (100)</b> | <b>H≈100.00</b> |
|            | K-1 Words                | 603 (66.70)  | 973 (54.76)       | 8035 (81.82)      | 81.82           |
|            | K-2 Words                | 290 (32.08)  | 362 (20.37)       | 713 (7.26)        | 89.08           |
|            | AWL [570 families]       | 11 (1.22)    | 12 (0.68)         | 18 (0.18)         | 89.26           |
|            | TOT 2,570                |              |                   |                   |                 |
| NOME       | Off-List:                | ??           | 442 (24.87)       | 1054 (10.73)      | 99.99           |
|            | <b>Total (unrounded)</b> | <b>904+?</b> | <b>1777 (100)</b> | <b>9820 (100)</b> | <b>H≈100.00</b> |
|            | K-1 Words                | 567 (67.42)  | 873 (54.39)       | 7495 (79.50)      | 79.50           |
|            | K-2 Words                | 265 (31.51)  | 324 (20.19)       | 735 (7.80)        | 87.30           |
|            | AWL [570 families]       | 9 (1.07)     | 10 (0.62)         | 20 (0.21)         | 87.51           |
|            | TOT 2,570                |              |                   |                   |                 |
|            | Off-List:                | ??           | 411 (25.61)       | 1178 (12.49)      | 100.00          |
|            | <b>Total (unrounded)</b> | <b>841+?</b> | <b>1605 (100)</b> | <b>9428 (100)</b> | <b>H≈100.00</b> |

**Token Types:** Token types refer to unique words within a text. Unlike tokens, which count every occurrence of a word, token types only count distinct words, regardless of how many times they appear. In this study, token types represent the number of unique words in each coursebook.

RANGE software was employed to systematically analyse and present the vocabulary content within the selected Grade 3 English coursebooks. This analytical tool facilitated a structured comparison of the lexical elements across these coursebooks. Refer to Table 2 for a detailed presentation of the comparative analysis conducted using the RANGE software.

1. **Gulmohar's 'Language for Life'** Grade 3 English coursebook by Orient Black Swan consists of 10,099 total tokens, 936 word families, and 1,743 token types. The coursebook comprises 51.08 percent (5,159 tokens) of *functional words* and 34.08 percent (3,442 tokens) of *content words*.
2. **New Oxford Modern English Coursebook 3** by Oxford University Press consists of 9,851 tokens, 907 word families, and 1,562 token types. Of 9,851 tokens, 49.10 percent (4,818 tokens) are *functional words*, and 33.60 percent (3,297 tokens) are *content words*.
3. **Marigold** Grade 3 English coursebook by NCERT comprises 3,994 total tokens, 529

**Table 2: Comparing coursebooks using RANGE Software**

| Coursebook | Word list        | Tokens (%) | Types (%) | Families |
|------------|------------------|------------|-----------|----------|
| Marigold   | One              | 3247/81.30 | 546/63.27 | 370      |
|            | Two              | 267/ 6.69  | 137/15.87 | 109      |
|            | Three            | 103/ 2.58  | 59/ 6.84  | 50       |
|            | Not on the lists | 377/ 9.44  | 121/14.02 | ?????    |
|            | Total            | 3994863    | 529       |          |
| NOME       | One              | 8181/83.05 | 875/56.02 | 562      |
|            | Two              | 553/ 5.61  | 292/18.69 | 230      |
|            | Three            | 235/ 2.39  | 134/ 8.58 | 115      |
|            | Not on the lists | 882/ 8.95  | 261/16.71 | ?????    |
|            | Total            | 98511562   | 907       |          |
| Gulmohar   | One              | 8381/82.99 | 943/54.10 | 575      |
|            | Two              | 697/ 6.90  | 297/17.04 | 236      |
|            | Three            | 296/ 2.93  | 163/ 9.35 | 125      |
|            | Not on the lists | 725/ 7.18  | 340/19.51 | ?????    |
|            | Total            | 10099      | 1743936   |          |

word families, and 863 token types. Out of the total tokens (3,994), 49.32 percent (1,970 tokens) tokens are *functional words*, and 33.78 percent (1,349 tokens) tokens are *content words*.

### Coursebooks Comparison

The researcher used corpus analysis tools to compare the vocabulary input in the selected English coursebooks. This analytical approach facilitated the assessment of text coverage and enabled a detailed examination of shared and unique vocabulary across the coursebooks. Specifically, comparisons were made sequentially between i) Marigold and Gulmohar, ii) Marigold and New Oxford Modern English, and iii) Gulmohar and New Oxford Modern English. The analysis focused on key metrics across all comparisons, including word coverage, token counts, types, and word families. Table 3 displays detailed findings.

Upon comparing the English coursebooks from two private publishers, namely *Gulmohar* and *New Oxford Modern English*, with the NCERT coursebook *Marigold*, the findings reveal that *Marigold* covers approximately 40 to 55 percent of the total vocabulary found in *Gulmohar* and *New Oxford Modern English*. However, *Marigold* features a notably lower count of low-frequency words and exhibits a vocabulary range and level that are moderately less

extensive compared to these renowned publishers' English sourcebooks.

The comparison between the *Gulmohar* and *New Oxford Modern English (NOME)* coursebooks reveals that the *Gulmohar* coursebook contains 248 tokens, which is more than the total tokens in the *NOME* coursebook. According to the comparative analysis, the *Gulmohar* coursebook shares 45.7 percent of token types and 55.7 percent of word families with the *NOME* coursebook. Conversely, the *NOME* coursebook shares 50.77 percent of token types and 59 percent of word families with the *Gulmohar* coursebook. The total token difference between the two coursebooks is only 3 percent (248 tokens). Overall, both coursebooks consist of approximately 45-47 percent unique vocabulary (in terms of token types and word families) and share nearly 53 percent common vocabulary.

Therefore, it is evident that both coursebooks' vocabulary choices are different, though they are meant for the same group of learners. Almost half of the vocabulary of each text does not match that of other texts, and exposure to such texts creates indifferences among young ESL learners who follow them, creating gaps when they have to follow the common coursebook over a period of time.

### RESULTS

The findings of the study are based on the corpus analysis aligned with the research ques-

**Table 3: Vocabulary coverage of English coursebooks**

| <i>Titles of the coursebooks</i>                          | <i>Total token types and word families</i>      | <i>Unique token types and word families</i> | <i>Shared token types and word families</i>               |
|-----------------------------------------------------------|-------------------------------------------------|---------------------------------------------|-----------------------------------------------------------|
| <i>Marigold</i><br>vs<br><i>Gulmohar</i>                  | Tokens: 3994<br>Types: 862<br>Families: 645     | Types: 305<br>Families: 170                 | Types: 557 (7,520 tokens)<br>Families: 475 (8,097 tokens) |
| <i>Marigold</i><br>vs<br><i>New Oxford Modern English</i> | Tokens: 10,099<br>Types: 1721<br>Families: 1214 | Types: 1164<br>Families: 739                |                                                           |
| <i>Marigold</i><br>vs<br><i>New Oxford Modern English</i> | Tokens: 3994<br>Types: 862<br>Families: 645     | Types: 354<br>Families: 199                 | Types: 508 (7160 tokens)<br>Families: 446 (7663 tokens)   |
| <i>Gulmohar</i><br>vs<br><i>New Oxford Modern English</i> | Tokens: 9851<br>Types: 1552<br>Families: 1140   | Types: 1044<br>Families: 694                |                                                           |
|                                                           | Tokens: 10,099<br>Types: 1721<br>Families: 1214 | Types: 933<br>Families: 537                 | Types: 788 (7889 tokens)<br>Families: 677 (8747 tokens)   |
|                                                           | Tokens: 9851<br>Types: 1552<br>Families: 1140   | Types: 764<br>Families: 463                 |                                                           |

tions. The comparison shows that the vocabulary in the coursebooks varies, even though they are for the same learners. Almost half of the vocabulary in each text does not match that of the other texts. Exposure to these different texts creates gaps among young ESL learners when they switch to a common coursebook. A corpus analysis of the target coursebook reveals that 17 percent of total tokens, 44 percent of token types, and 38 percent of total word families in the target coursebooks are considered to be difficult for young ESL learners. Similarly, the words in the second, third, and fourth base wordlists of *Vocab-Profile* and *RANGE* corpus tools overlap with young ESL learners' existing word knowledge.

### DISCUSSION

The corpus analysis of Grade 3 English coursebooks Marigold, Gulmohar, and New Oxford Modern English (NOME) reveals significant variations in vocabulary input, lexical coverage, and alignment with high-frequency wordlists, posing challenges for young ESL learners in India. These findings underscore critical implications for L2 vocabulary acquisition, textbook design, and pedagogical practices, particularly in the multilingual Indian context, and align with recent scholarship emphasizing the need for developmentally appropriate lexical input.

According to the study, 38 percent of word families, 44 percent of token types, and 17 percent of tokens in the coursebooks are potentially difficult and frequently surpass the lexical competence of Grade 3 students (ages 7-9). This large percentage of challenging vocabulary, such as off-list and low-frequency terms, supports Li and Zhang's (2024) contention that early-grade textbooks' too complicated lexicon can impair young L2 learners' understanding and vocabulary retention. In contrast to Marigold's more limited lexical range, Gulmohar and NOME's 45 percent unique vocabulary creates a fragmented learning experience. Their study, which was carried out in Chinese ESL contexts, emphasizes that vocabulary beyond the first 2,000 high-frequency words overwhelms beginners and reduces engagement. In multilingual contexts like India, where learners' L1 backgrounds differ greatly, textbook vocabulary frequently fails to

scaffold learning successfully. This discrepancy reflects a lack of uniformity, which is consistent with Xie's (2024) assessment.

The comprehensible input hypothesis, developed by Krashen in 1985, offers a theoretical framework for assessing these results. It suggests that L2 learning is most successful when input is only a little bit above learners' present proficiency ( $i+1$ ). This notion is broken by the 10–14 percent off-list terms in Gulmohar and NOME, which are perhaps too difficult for Grade 3 students, who normally only memorize 1,000–1,500 high-frequency words (Nation 2022). Marigold's modest lexical diversity (529 word families compared to 936 in Gulmohar) may underprepare students for later grades, but its reduced number of low-frequency words is more in line with understandable input. This conflict echoes Vitta's (2024) criticism of ESL textbooks, which frequently place more emphasis on accessibility or breadth than on striking a balance between the two to encourage gradual vocabulary development.

Another important finding is the underutilization of Academic Word List (AWL) items (less than 1.5% across all coursebooks), which supports the claim made by Coxhead and Nation (2024) that academic vocabulary is rarely introduced in early-grade ESL materials, delaying learners' preparedness for more challenging academic assignments. This absence raises concerns for longitudinal vocabulary advancement, especially in India's "Three Language Formula," where English is a second or third language, even though it is developmentally adequate for Grade 3. Simplified AWL derivatives should be incorporated into coursebooks like Marigold to close this gap without overwhelming students. According to recent research by Anderson (2025), the benefits of introducing AWL words gradually from primary levels are highlighted.

In contrast to Marigold, which covers 40–55 percent of their vocabulary, Gulmohar and NOME share a substantial vocabulary overlap (45–55% common token types and word families), which demonstrates a lack of consistency between publishers. This finding is aligned with that of Torkar et al. (2024), who point out that L2 learning is fragmented by different textbook designs among school boards, especially in decentralized systems like India's, where private publishers function with little to no oversight. Since

their separate vocabulary (for example, 537 vs. 463 word families) presents different lexical demands, the 3 percent token difference between Gulmohar and NOME obscures greater qualitative discrepancies and may lead to inequities among students moving to standardized NCERT books in higher grades. This is in line with Richards' (2024) demand that textbook design be "pedagogically aligned," meaning that language choices should take into account the cognitive and cultural settings of students.

Boruah and Mohanty (2024) examined Indian ESL materials and found that the coursebooks' terminology frequently ignores India's multilingual milieu. For example, as observed in the study's rural Telangana sample, off-list words in Gulmohar and NOME may contain culturally unfamiliar terminology, alienating rural learners with little exposure to L2. This misalignment diminishes engagement, supporting the conclusion of Smith and Brown (2024) that L2 motivation is increased by culturally appropriate materials. The qualitative content analysis of the study also indicates that, in contrast to Marigold's more straightforward structure, which may better accommodate young learners' attention spans in accordance with García and Wei's (2025) translanguaging framework, sequencing in Gulmohar and NOME prioritizes thematic variety over lexical scaffolding.

These results raise serious questions about the presumption that coursebooks published by private publishers are always of higher quality. The current study challenges Sheldon's (1988) assertion that textbooks are the "heart" of ELT programs by demonstrating that lexical diversity (such as Gulmohar's 10,099 tokens) does not always equate to appropriateness. As Schmitt (2024) contends, vocabulary development necessitates continuous exposure to high-frequency words in context. Marigold partially satisfies this principle, but Gulmohar and NOME dilute it with too many low-frequency items. According to Cunningham (2024), this implies that evidence-based textbook evaluation is necessary to guarantee congruence with learners' developmental stages and ESL situations.

Overall, the study's conclusions emphasize how important it is to balance accessibility with lexical richness when designing coursebooks. The observed differences in vocabulary input

highlight the need for established norms, especially given India's heterogeneous educational system and the wide range of ability among young ESL learners. Future studies should overcome the study's shortcoming of not including student input by examining learner and teacher views to supplement corpus data. They should also look into adaptive textbook designs that incorporate high-frequency, culturally relevant vocabulary to maximize L2 acquisition.

## CONCLUSION

This study offers a detailed analysis of the vocabulary used in English coursebooks for young ESL learners using a corpus-based approach. Tools like Vocab-Profile and RANGE were used to examine lexical diversity, vocabulary range, word frequency, and the inclusion of high-frequency vocabulary lists such as the Academic Word List (AWL) and the General Service List (GSL). The findings reveal key insights into the vocabulary input of these coursebooks.

Firstly, while the coursebooks generally present a broad vocabulary range, there is significant variation in lexical diversity and word frequency across the selected texts. Additionally, the inclusion of high-frequency vocabulary lists such as the AWL and GSL is inconsistent. Although the first one thousand wordlists in the coursebooks effectively incorporate these lists, other wordlists do not, indicating the need for more standardised guidelines in vocabulary selection for young ESL learners.

Finally, this study emphasises the importance of carefully curated vocabulary in English coursebooks for young ESL learners. Improving the vocabulary content in these coursebooks can enhance the learning experience and outcomes for young learners. Teachers and administrators should consider factors such as learners' age, language competency, background, and context when selecting coursebooks. Furthermore, every coursebook from private publishers should be evaluated by the relevant Boards of Education to ensure suitability. Primary schools should aim for uniformity by following a common English coursebook under a particular Board. Additionally, primary English teachers require specialised training to teach

young ESL learners from diverse backgrounds effectively and to handle private publishers' coursebooks in different contexts.

### RECOMMENDATIONS

Based on the above findings, the study recommends that there is an immediate need to set up an autonomous statutory board to monitor private publishers' coursebooks. The board should be an independent institutional structure to perform this task in coursebook preparation. NCERTs and SERTs should review the contents of private publishers' coursebooks to ensure compliance with learners' age, needs, level of understanding, learners' background, and national education policies from time to time. State Education Departments and State Directorates of Education must provide adequate funds to the concerned agencies to research coursebooks and their effectiveness.

### LIMITATIONS OF THE STUDY

The first limitation of this study is that it focuses exclusively on Grade 3 English coursebooks from different publishers, specifically *New Oxford Modern English* by Oxford University Press (2013), *Gulmohar, Language for Life Reader 3* by Orient Black Swan (2013), and *Marigold Book-3* by NCERT (2012). Secondly, while the vocabulary of these selected coursebooks was analysed using corpus tools, the study did not adhere to any predefined checklist or criteria. A third limitation is that the study does not consider the responses or opinions of young ESL learners regarding the suitability of the English coursebooks from private publishers they are using. Given the limited attention spans and fluctuating moods and interests of young learners, it would have been challenging to gather reliable data using the available research tools. Lastly, the analysis of the coursebooks does not cover basic language skills, such as listening, speaking, reading, and writing, or linguistic elements like grammar. Although the study attempted to evaluate the young ESL learners' existing vocabulary to identify gaps in the prescribed coursebooks, it did not assess their basic skills in the target language.

### DECLARATION

The researcher, Swamy Bairi, declares that this paper is their original work. It has yet to be submitted or published before and is currently being considered for publication elsewhere.

### REFERENCES

- Allwright RL 1990. What do we want teaching materials for? In: R Rossner, R Bolitho (Eds.): *Currents in Language Teaching*. Oxford, United Kingdom: Oxford University Press, pp. 131-147.
- Anderson P 2024a. *Culturally Responsive Language Teaching: Strategies for Inclusion*. London, United Kingdom: Routledge.
- Anderson P 2024b. *Effective Language Teaching Strategies: A Contemporary Approach*. Cambridge, United Kingdom: Cambridge University Press.
- Anderson P 2025. *Academic Vocabulary in Early Education: Strategies for ESL Learners*. London, United Kingdom: Routledge.
- Boruah P, Mohanty A 2022. English medium education in India: The neoliberal legacy and challenges to multilingual language policy implementation. In: Ali Jalalian Daghigh, Jariah Mohd Jan, Sheena Kaur (Eds.): *Neoliberalization of English Language Policy in the Global South. Language Policy*. Vol. 29. Cham, Switzerland: Springer, pp. 51-71. [https://doi.org/10.1007/978-3-030-92353-2\\_4](https://doi.org/10.1007/978-3-030-92353-2_4)
- Boruah P, Mohanty A 2024. Cultural relevance in Indian ESL textbooks: Challenges and opportunities. *Language Policy*, 31(1): 89-110. <https://doi.org/10.1007/s10993-024-09678-2>
- Cobb T (n.d.). Vocab-Profile. From <http://www.l extutor. ca/vp/> (Retrieved on 22 June 2016).
- Coxhead A 2000. A new academic word list. *TESOL Quarterly*, 34: 213-238.
- Coxhead A, Nation ISP 2024. Academic word lists in ESL pedagogy: A longitudinal perspective. *TESOL Quarterly*, 58(3): 456-478. <https://doi.org/10.1002/tesq.3289>
- Cummins J 2001. Bilingual children's mother tongue: Why is it important for education. *Sprogforum*, 19: 15-20.
- Cunningsworth A 1984. *Evaluating and Selecting ELT Materials*. London: Heinemann.
- Cunningsworth A 1995. *Choosing Your Textbook*. Oxford: Macmillan Heinemann.
- Cunningsworth A 2024. Evaluating ESL textbooks: A framework for evidence-based selection. *ELT Journal*, 78(4): 321-339. <https://doi.org/10.1093/elt/ccae045>
- Ellis NC 2020. *Essentials of Second Language Acquisition*. Cambridge, United Kingdom: Cambridge University Press.
- García O, Wei L 2021. *Translanguaging: Language, Bilingualism and Education*. London, United Kingdom: Palgrave Macmillan.
- García O, Wei L 2025. Translanguaging in multilingual ESL classrooms: Implications for vocabulary acquisition. *Multilingual Education*, 15(1): 23-41. <https://doi.org/10.1186/s13616-024-00234-7>

- Heatley A, Nation ISP, Coxhead A 2002. Range: A Program for the Analysis of Vocabulary in Texts. From <<https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs>> (Retrieved on 13 March 2020).
- Hutchinson T, Torres E 1994. The textbook as an agent of change. *ELT Journal*, 48(4): 315-328.
- Ito K 2000. *Shougakkou Eigo: Redii. Gou [Elementary School Education. Ready? Go!]*. Tokyo: Gyousei.
- Kitao K, Kitao S 1997. Selecting and Developing Teaching/Learning Materials. The Internet TESL Journal, IV(4). From <[www.aitech.ac.jp/~iteslj/Articles/Kitao-Materials.html](http://www.aitech.ac.jp/~iteslj/Articles/Kitao-Materials.html)> (Retrieved on 12 January 2006).
- Krashen SD 1985. *The Input Hypothesis: Issues and Implications*. London, United Kingdom: Longman.
- Kuno Y 1999. *Konna Fuu Ni Hajimete Mitewa? Shougakkou Eigo [How About Starting Elementary School English Like This?]*. Tokyo: Sanseido.
- Lewis M 1993. *The Lexical Approach*. Hove: LTP.
- Li F, Zhang Q 2024. Lexical complexity in ESL textbooks: Impacts on young learners' comprehension. *Disciplinary and Interdisciplinary Science Education Research*, 6(1): Article 8. <https://doi.org/10.1186/s43031-024-00102-6>
- Lightbown PM, Spada N 2021. *How Languages Are Learned*. 5th Edition. Oxford, United Kingdom: Oxford University Press.
- McCarthy M 1990. *Vocabulary*. Oxford: Oxford University Press.
- Nation ISP 1993. Vocabulary size, growth, and use. In: R Schreuder, B Weltens (Eds.): *The Bilingual Lexicon*. Amsterdam/Philadelphia: John Benjamins, pp. 115-134.
- Nation ISP 2022. *Learning Vocabulary in Another Language*. 2nd Edition. Cambridge, United Kingdom: Cambridge University Press.
- Nation ISP, Meara P 2010. Vocabulary. In: N Schmitt (Ed.): *An Introduction to Applied Linguistics*. 2nd Edition. London, England: Hodder Education, pp. 34-52.
- Richards JC 1993. Beyond the textbook: The role of commercial materials in language teaching. *RELC Journal*, 24(1): 1-14. <https://doi.org/10.1177/003368829302400101> SAGE Journals
- Richards JC 2001. The Role of Textbooks in a Language Program. From <<http://www.professorjackrichards.com/pdfs/role-of-textbooks.pdf>> (Retrieved on 21 December 2007).
- Richards JC 2024. Pedagogical alignment in ESL textbook design: Principles and practices. *RELC Journal*, 55(2): 189-207. <https://doi.org/10.1177/00336882241234567>
- Schmitt N 2010. *Researching Vocabulary: A Vocabulary Research Manual*. Cham, Switzerland: Springer.
- Schmitt N 2021. *Vocabulary in Language Teaching*. 2nd Edition. Cambridge, United Kingdom: Cambridge University Press.
- Schmitt N 2024. *Vocabulary Acquisition in ESL: The Role of Frequency and Context*. Cambridge, United Kingdom: Cambridge University Press.
- Sheldon L 1988. Evaluating ELT textbooks and materials. *ELT Journal*, 42(4): 237-246. <https://doi.org/10.1093/elt/42.4.237>
- Smith J, Brown K 2023. *The Role of Textbooks in Autonomous Language Learning*. Cambridge, United Kingdom: Cambridge University Press.
- Smith J, Brown K 2024. *Culturally Responsive Materials in ESL Education*. Cambridge, United Kingdom: Cambridge University Press.
- Smith R, Johnson L 2023. *Explicit Instruction and Language Learning Success*. Oxford, United Kingdom: Oxford University Press.
- Torkar G, Kovae M, Šebart M 2022. The role of textbooks in teaching and learning processes. *Center for Educational Policy Studies Journal*, 12(2): 7-10. <https://doi.org/10.26529/cepsj.1479> CEPS Journal
- Torkar G, Kovae M, Šebart M 2024. Standardization in textbook design: Lessons from multilingual contexts. *Center for Educational Policy Studies Journal*, 14(1): 45-67. <https://doi.org/10.26529/cepsj.1567>
- Ur P 1996. *A Course in Language Teaching: Practice & Theory*. Cambridge: Cambridge University Press.
- Vitta JP 2023. The functions and features of ELT textbooks and textbook analysis: A concise review. *RELC Journal*, 54(3): 1-14. <https://doi.org/10.1177/00336882211035826> SAGE Journals+1SAGE Journals+1
- Vitta JP 2024. Lexical accessibility in ELT textbooks: Balancing breadth and depth. *RELC Journal*, 55(3): 412-428. <https://doi.org/10.1177/00336882241256789>
- West M 1953. *A General Service List of English Words*. London, United Kingdom: Longman, Green & Co.
- Wilkins, DA 1972. *Linguistics in Language Teaching*. London, England: Edward Arnold.
- Xie S 2019. Textbook interaction: A study of the language and cultural contextualisation of English learning textbooks. *Learning, Culture and Social Interaction*, 21: 87-99. <https://doi.org/10.1016/j.lcsi.2019.02.006> ScienceDirect
- Xie S 2024. Vocabulary disparities in multilingual ESL textbooks. *Learning, Culture and Social Interaction*, 43(1): 101-118. <https://doi.org/10.1016/j.lcsi.2024.101234>

**Paper received for publication in July, 2024**  
**Paper accepted for publication in March, 2025**