



SYNTACTIC COMPLEXITY IN INTERMEDIATE ENGLISH TEXTBOOK 1: A STUDY OF SENTENCE STRUCTURE AND CEFR COMPLIANCE

Azhar Munir Bhatti
(Corresponding Author)

PhD Scholar, University of Education, Lahore

Email: azharmunir18@gmail.com

Prof. Dr. Ahsan Bashir

Chairman Department of English, University of Education, Lahore

Abstract

This paper explores the syntactic complexity of Intermediate English Textbook 1, a textbook used as a core subject at the intermediate level of learners of the English language in Pakistan, and the purpose of the study is to assess whether the text satisfies the Common European Framework of Reference to Languages (CEFR). The study utilized a corpus-based quantitative design, as it compared sentence- and clause-level structures, types of phrases, and distributions of tenses/aspects in the two halves of the textbook. The Mean Length of Sentence (MLS), Mean Length of Clause (MLC) and Clauses per Sentence (C/S) were calculated with the L2 Syntactic Complexity Analyzer and compared with reference values of CEFR B1-B2. Findings showed that the general syntactic picture fits the B1-B2 proficiency band - the even distribution of clauses and moderate subordination, however, the syntactic complexity development across units is not extensive. The text is structurally sufficient with no developmental escalation, which is what the recent scholarship terms as the absence of developmental complexity. According to the qualitative interpretation, a lack of gradation in the clause embedding, phrase elaboration, and perfect forms and conditional forms limits the progress of learners to CEFR-B2 autonomy. The paper ends with a conclusion that ELT materials used in Pakistan need to be organized in terms of syntactic advancement and empirically based evaluation indices to make sure that they are in line with international proficiency standards. Pedagogical implications involve the redesign of textbooks with syntactic scaffolding levels which are graded, text genres diversification and incorporation of C1-level structures to promote structural fluency and communicative sophistication of learners.

Keywords: *Syntactic complexity; CEFR alignment; Textbook evaluation; Clause and phrase analysis; Pakistani EFL; Corpus-based study; Intermediate proficiency*

1. INTRODUCTION

The recent tendency to match English-language learning materials with the Common European Framework of Reference on Languages (CEFR) has predetermined the necessity to study not only lexical and discourse features of textbooks, but also the syntactic complexity of texts that they offer. It has been proposed that the syntactic complexity, which is usually measured using indicators like mean length of sentence, number of clauses in a T-unit, number of dependent clauses in a clause, etc., is associated with the level of the learner and the complexity of the text (e.g., Zhang, 2024; Li et al., 2025). Further, the textbook materials, which do not raise the structural complexity in a graded manner, might not correspond to the anticipated stages of the CEFR levels and, thereby, hinder the scaffolding of syntactic development by learners (Zhang & Lu, 2022; Issa



et al., 2022). Within the context of Pakistan and other EFL environments, the textbook is currently the most dominant input to the learners; however, research has revealed that intermediate level textbooks might have a great level of lexical or grammatical load but lack the corresponding syntactic development, thus making it more difficult to help the learners map form to meaning and achieve alignment with CEFR descriptors (Issa et al., 2022; Noor et al., 2025). Since the syntactic competence is the core of writing and speaking in the higher levels, it is high time to have the Intermediate English Textbook 1 undergoing a syntactic complexity audit, and measuring its sentence-structure profile, as well as its adherence to the syntactic requirements that are associated with the CEFR.

A transitional textbook must serve as an intermediary between the lower-order types of grammatical processing and the autonomous syntactic control of B2 or C1 CEFR levels in the transition between lexically dense or structurally simple materials and materials that increasingly mobilise complex clause-embedding and phrase-level sophistication (Saricaoglu & Atak, 2022; Khushik, 2020). Empirical studies of learner writing emphasize that syntactic complexity depends on their level of proficiency: more advanced speakers use more dependent clauses and noun-phrase modifiers, which are more indicative of high levels of syntactic maturity (Saricaoglu & Atak, 2022). There would be an extension of this argument on textbook resources where Intermediate English Textbook 1 would provide no graded exposure to these structures and in turn, unwillingly limit the readiness of learners to more advanced tasks. The question of structural analysis in addition to lexical features is indicated in the work of Bhatti on textbook vocabulary and ELT materials (Mushtaq & Bhatti, 2021) and overall criticism of the textbook structure (Asif, Saeed & Kang, 2021) in Pakistan. This paper, therefore, presents the thesis that the textbook should be analyzed using syntactic complexity as an indicator to not only determine whether the sentence-structure profile of the textbook is compatible with incremental needs of the CEFR but to also allow a consistent syntactic development in the intermediate learners.

1.1. Research Questions

- What is the overall level of syntactic complexity in *Intermediate English Textbook 1* as measured through clause-based and phrase-based indices?
- How does the syntactic complexity of the textbook's reading passages and exercises align with CEFR-based descriptors for B1–B2 levels?
- To what extent does syntactic complexity progress across units, reflecting graded difficulty in line with CEFR expectations?

2. LITERATURE REVIEW

The past ten years have rejuvenated interest in the role of syntactic complexity (SC) indexes of text difficulty and proficiency, and in whether the curricular material is CEFR-aligned. Research on L2 writing repeatedly demonstrates that the increases of clausal subordination, phrasal elaboration, and the length of the sentences in general correlate with the improvement in proficiency, although they also warn that the indices do not always behave similarly across the genres and tasks (e.g., Atak, 2021; Saricaoglu & Atak, 2022; Zhang, 2024). To apply it to textbook texts, Zhang and Lu (2022) suggest matching measurable linguistic characteristics with CEFR bands so that the input difficulty increases in line with the desired levels, which have recently been



reflected by research that pairs SC measures with readability models in ELT textbook series (Li et al., 2025). At the scale of assessment systems, checklist study reveals that a lot of the CEFR-branded resources do not have a clear, empirically supported complexity gradation, especially at the sentence-structure level, which may indicate the necessity to involve SC diagnostics in the daily textbook evaluation.

The recent analysis in the EFL context of Pakistan where the textbook is the main input is a mixed picture. A quantitative analysis of intermediate level English textbook yielded disproportional profiles of linguistic complexity of selections, which demonstrate the question of progression and teachability (Issa, Hussain, & Abbas, 2022). Similar work on the readability of HSSC materials also reports an inconsistency between texts of the same level, which would mean that the results are not matched with the linguistic requirements (Hussain, 2022). The British Council system reviews further add that, although policy is increasingly making reference to CEFR, classroom resources, and assessment practices have not been consistent in operationalizing CEFR-aligned linguistic targets. In textbook-internal aspects, lexical (e.g., corpus-based vocabulary research of Intermediate Book 1) has been by far the most empirically studied area, and sentence-level structure has remained comparatively under-investigated (Mushtaq, Bhatti, & Yasmin, 2021). In addition to Pakistan, a more recent systematic review of CEFR-aligned textbooks also observes heterogeneous construct validity in the demonstration of difference between the evidences of CEFR-aligned, which underlines the argument of feature-based audits that incorporate SC (Hamid et al., 2025).

Even with these developments, we are yet to have a textbook-specific, sentence-structure audit, which (a) characterizes the syntactic complexity of Intermediate English Textbook 1 in units and genres, and (b) assesses its adherence to the CEFR B1-B2 expectations through validated SC indices (e.g., clauses/T-unit, DC/C, NP-modifiers), relative to CEFR-anchored standards. Available Pakistani literature either lacks the isolation of linguistic complexity as well as the isolation of SC, emphasizes the concept of readability, or is more concerned with the lexical measure; none of them systematically relates sentence-level characteristics to the thresholds of the CEFR referenced in this particular textbook (Issa et al., 2022; Hussain, 2022; Mushtaq et al., 2021). At the same time, foreign assignments offer powerful approaches to text classification based on CEFR and to correlating SC with difficulty, but has seldom been tailored to intermediate text of Pakistan (Zhang & Lu, 2022; Li et al., 2025). To fill this gap, the current paper proposes a fine-grained SC analysis of Intermediate English Textbook 1, which would combine CEFR-based logic of alignment with measurable SC parameters to inform evidence-based textbook revision and teaching. (Zhang, 2024; British Council, 2022).

3. METHODOLOGY

3.1. Research Design

The paper follows a quantitative corpus-analytic design, which is enriched with a comparative interpretive framework that is in line with the CEFR descriptors. The aim is to assess the syntactic complexity of Intermediate English Textbook 1 which is taught in intermediate colleges in Punjab, Pakistan. The methodology is based on the previous CEFR-connected textbook research (Zhang & Lu, 2022; Li et al., 2025) but adapts its steps to the Pakistani EFL resources. The study



incorporates computational text analysis of objective syntactic metrics alongside qualitative analysis of patterns in comparison to the CEFR-defined proficiency standards (Issa et al., 2022; Mushtaq & Bhatti, 2021).

3.2. Corpus and Sampling

The corpus includes all reading passages, dialogues, comprehension texts, and writing-model of Intermediate English Textbook 1 (Punjab Curriculum and Textbook Board, 2022 edition). There was no use of grammar exercises, vocabulary lists, or poems to make sure that the data used was a representative of real continuous prose. The corpus was scanned and purged of formatting errors and divided into units and text type (narrative, descriptive, expository). The segments were used as sub-corpus to compare themselves internally and therefore, the syntactic development of the textbook could be analyzed.

3.3. Analytical Framework and Tools

Syntactic complexity was measured using a combination of clause-based, T-unit, and phrase-level indices recommended in second language complexity research (Lu, 2011; Kyle, 2020). Key indices include:

- Mean length of sentence (MLS)
- Clauses per T-unit (C/T)
- Dependent clauses per clause (DC/C)
- Complex T-units per T-unit (CT/T)
- Coordinate phrases per clause (CP/C)
- Noun phrase modifiers per nominal (NP-Mod/N)

The study employed L2 Syntactic Complexity Analyzer (L2SCA) and Text Inspector (CEFR Edition) to extract these features automatically. Statistical averages and standard deviations were calculated for each index across units. To evaluate CEFR alignment, syntactic scores were compared with CEFR B1–B2 reference benchmarks derived from Zhang and Lu (2022) and Li et al. (2025), where approximate ranges for each index are provided based on large learner corpora.

3.4. Data Interpretation

Triangulation of quantitative findings with qualitative content analysis was done to explain whether incremental learning can be supported by sentence structures. The specific focus was placed on the patterns of clause-embedding, application of subordination and distribution of complex noun phrases- the important indicators of CEFR level sophistication. Pedagogical mismatch (e.g., sudden changes in SC in successive units) was coded as non-CEFR compliant. It is hoped that the results will shed light on whether the syntactic requirements of the textbook are relevant to the target level of proficiency of the learners or not.

3.5. Ethical Considerations and Reliability

No human subjects were used since the source of data is publicly available instructional material. This was done by ensuring analytical reliability through repeated corpus validation, cross-checking of indices across tools and inter-rater agreement during the manual verification phase ($k = 0.92$). Each quantitative process was repeated twice in order to obtain consistency of extracted values.



4. RESULTS

Table 1

Overall Syntactic Complexity of Intermediate English Textbook 1

Measure	Book 1 (1st Half)	Book 1 (Last Half)	Combined Average	CEFR Reference Range (B1– B2)*	Interpretation
Number of Sentences	1116	996	2112	–	Sufficient sentence volume for intermediate corpus; balanced unit lengths.
Total Clauses	1650	1470	3120	–	Consistent clause density across halves.
Clauses per Sentence (C/S)	1.48	1.48	1.48	1.40–1.60	Typical for B1–B2 texts, reflecting moderate subordination.
Mean Length of Sentence (MLS)	12.34	12.68	12.51	11–13	Aligned with CEFR intermediate target.
Mean Length of Clause (MLC)	8.36	8.59	8.48	8–10	Appropriate for intermediate discourse.
Dependent Clauses %	32 %	31 %	31.5 %	30–35 %	Balanced mix of simple and complex structures.
CEFR Grammar Measure	B1–B2 (C1 elements)	B1–B2 (C1 elements)	B1–B2	–	Reflects graded but stable complexity across halves.

*CEFR reference values adapted from Zhang & Lu (2022) and Li et al. (2025).

The overall syntactic profile of *Intermediate English Textbook 1* remains stable across halves, showing a balanced distribution of simple and complex sentences consistent with CEFR B1–B2 levels. Slightly higher MLC and MLS in the second half indicate incremental difficulty, aligning with the CEFR requirement for narrative elaboration and complex event sequencing. The 31–32 % dependent-clause ratio suggests that learners are exposed to moderately nested structures—enough to challenge but not overwhelm intermediate users.

Table 2

Clause-Type Distribution in Book 1

Clause Type	1st Half (Freq.)	Last Half (Freq.)	% Change	Pedagogical Function
Independent Clauses	1116	996	–10.8 %	Maintains narrative flow; supports main ideas.



Adverbial Clauses	350	300	–14.3 %	Temporal & conditional relations.
Relative Clauses	130	120	–7.7 %	Descriptive subordination; typical of B2.
Nominal Clauses	40	35	–12.5 %	Argument complement structures.
Complement Clauses	14	19	+35.7 %	Verb-complement expansion at higher units.

Clause analysis reveals that dependent clause density slightly decreases overall, except for complement clauses, which grow in frequency toward later units. This increase aligns with C1-edge syntactic behavior, such as verbs of cognition (“believe that,” “expect to”), promoting abstract expression. The reduction of adverbial and relative clauses may reflect shorter moral-parable texts in later chapters, yet the complement-clause growth indicates some deliberate grammatical enrichment.

Table 3

Phrase-Type Distribution

Phrase Type	1st Half (Freq.)	Last Half (Freq.)	Combined Total	CEFR Alignment Comment
Noun Phrases (NP)	2100	2000	4100	High density supports lexical complexity development.
Verb Phrases (VP)	1200	1100	2300	Balanced use of simple and phrasal verbs.
Prepositional Phrases (PP)	1000	950	1950	Typical for B1–B2; adds spatial and temporal detail.
Adjective Phrases (AdjP)	150	100	250	Moderate use – aligns with descriptive focus.
Adverb Phrases (AdvP)	100	50	150	Slight decline in later units – possible style simplification.

Phrase-structure analysis demonstrates noun-phrase prominence, reinforcing CEFR’s emphasis on *information packaging* at upper-intermediate levels. The consistent VP and PP frequencies indicate steady syntactic scaffolding. However, the decline in adjectival and adverbial phrase density may imply a narrative simplification in later lessons, suggesting less exposure to elaborative modification—an area textbook designers might strengthen for fuller CEFR compliance.

Table 4

Tense, Aspect, and Modal Distribution

Grammatical Category	1st Half (Freq.)	Last Half (Freq.)	Combined %	CEFR Interpretation
Present Simple	400	350	22 %	Used for statements and truths; core B1.



Past Simple	600	550	34 %	Dominant for narration; B1–B2 indicator.
Present Perfect	30	25	2 %	Emergent C1 feature but under-represented.
Past Perfect	40	30	2 %	C1 indicator of event sequencing; needs reinforcement.
Progressive Forms	50	40	2 %	Moderate use – supports discourse variety.
Modals	200	150	11 %	Functional range adequate for B2 communicative tasks.
Conditionals	25	20	1.3 %	Limited exposure; scope for greater CEFR alignment.

Verb-system distribution underscores a past-tense dominance, suitable for narrative progression but showing underrepresentation of perfect and conditional forms that typify B2–C1 progression. While modals are adequately frequent, expanding the range of perfect and hypothetical constructions could better serve CEFR communicative competence goals (e.g., “would have gone,” “might have been”).

Table 5

CEFR Alignment Summary

Dimension	Observed Level	CEFR Range	Observed Features	Pedagogical Implication
Syntactic Complexity (MLS, MLC, C/S)	B1–B2 (Stable progression)	B1→B2	Moderate clause embedding, few C1 structures	Aligned with CEFR; requires more graded complexity.
Phrase-Level Development	Upper B1	B1–B2	High NP density; low AdjP/AdvP use	Enhance modification and description.
Verb System Range	B1	B1–C1	Limited perfect & conditional forms	Need more advanced tense variety.

The synthesized information proves that Intermediate English Textbook 1 has linguistic competency to intermediate skills, but the syntactic development is not developmental, but linear. Although the clauses and phrases will be constant, it is possible that the absence of gradual complexity between halves will decrease exposure to syntactic escalation, which is a hallmark of CEFR. It is advised that designers include more multi-clausal and modifier-rich constructions in order to recreate natural B2-C1 text conditions.

5. DISCUSSION

The syntactic analysis of Intermediate English Textbook 1 shows that the syntactic profile is B1-B2 with occasional C1-level elements, which proves that the syntactic analysis is partially consistent with the expectations of intermediate learners according to the CEFR. Quantitatively,



the measures of Clauses per Sentence (1.48), Mean Length of Sentence (12.5), and Mean Length of Clause (8.4) are quantitatively in the intermediate range of the CEFR (Zhang & Lu, 2022). The results are consistent with the previous research showing that syntactic indices such as the density of clauses and the length of sentences are reliable indicators of CEFR progression when incorporated into the textbooks analysis (Li et al., 2025; Kyle, 2020). Nevertheless, the numerical profile is the same, but the absence of the syntactic increase by half implies that in the text, the similar grammatical structures could be repeated without enough scaffolding to support the advanced learners. As Asif, Saeed and Kang, (2021) warns, the design of textbooks in Pakistan tends to favor thematic over structural gradation, resulting in stylistically enriched but syntactically stagnant materials that constrain the developmental path of the learners.

This state can be highlighted by having a closer look at the patterns of clauses and phrases. The dependent-clause ratios (31 %) and equal noun-phrase densities show that it was a plateau of the syntactic and not a pedagogical ascent. This validates previous Pakistani results that language complexity in national syllabi is usually the same at different levels (Issa et al., 2022; Noor et al., 2025). In terms of CEFR, the true B2 development requires the development of more subordinate clauses embedding and nominal-phrase elaboration- aspects that can only be partially seen here. Recent corpus-based studies associate this syntactic densification with cognitive and communicative maturity (Saricaoglu & Atak, 2022), which consequently leads to the development of academic-register. The use of narrative forms instead of expository or argumentative types of text in the textbook can also be the cause of poor syntactic diversification (Mushtaq & Bhatti, 2021). Therefore, although the book has the grammatical transparency needed to be understood, it fails to provide learners with the syntactic challenge, which is an important CEFR indicator of the transition between controlled accuracy and fluent flexibility.

The paper therefore reveals a pedagogical fault: Intermediate English Textbook 1 attains surface adherence to CEFR descriptors but fails to advance through syntactic escalation. Its syntactic profile resembles what Zhang (2024) refers to as a form of static complexity grammatical texts that are correct but pedagogically stagnant as they do not take learners out of familiar structures. Although the ACTFL classification decides to classify the material as Advanced Low, the CEFR needs to have visible progression between B1 and B2 by increasing complexity stepwise in the structure of the clauses and in the phrasal modification. The absence of perfect and conditional forms and the limited application of adjectival/adverbial phrases are indicative of the lack of the coverage of the higher-order syntactic categories. Similar to the critique of Pakistani textbook linguistics by Bhatti (2021), this paper holds that data-driven revision is necessary: by incorporating multi-clausal tasks, raising the proportion of relative clauses, and adding a variety of text types that encourage the development of syntactic variation. This reform would not only align the national materials to the expectations of the CEFR, but also guarantee the intermediate learners to internalize the structural flexibility that is inherent in effective use of the English language.

6. CONCLUSION

This paper aimed to assess the syntactic complexity of Intermediate English Textbook 1 with respect to its mapping onto the Common European Framework of Reference in Languages (CEFR)



in the Pakistani intermediate EFL situation. The findings showed that even though the textbook has the syntactic features typical of a B1-B2 profile (mean length of sentences, number of clauses per sentence, and frequency of dependent clauses), the gradual increase in the complexity of the units is quite low and does not include the well-defined progressive levels of scaffolding of the CEFR requirements. This observation is in line with recent studies that reveal that a lot of ELT resources meet the requirements of surface grammar only to fail to promote structural growth in the long term (Zhang & Lu, 2025; Anggia & Habok, 2023). Specifically, such a low range of adjectival/adverbial phrases density and the under-representation of the forms of perfect and conditional verbs indicate that, perhaps, the learners are not sufficiently equipped to address the cognitive and communicative challenges of upper-intermediate assignments. Therefore, we can conclude that the textbook partially fulfills the requirements of CEFR, but it does not exhaust the usage of syntactic complexity as the pedagogical tool of the learner development.

Pedagogically and materially-developmentally speaking, several crucial implications arise. To begin with, textbook designers and syllabus planners are advised to incorporate explicit syntactic escalation routes, such as intentional increments in multiple-clause embedding, extension of nominal modifier chains, and systematic additions of C1-type structures, to make sure that learners no longer focus on accuracy but on linguistic flexibility (Golparvar, Casal & Abolhasani, 2025; Bele, 2025). Second, the textbook should be used by TEFL teachers in Pakistan as a starting point, with the extra texts or exercises that intentionally bring the learners to more complicated structures than the provided ones, thus filling the gap between the textbook plateau and the upward horizon of the CEFR. Third, instead of using thematic coherence or lexical coverage, curriculum authorities need to use empirical complexity diagnostics (like syntactic indices and readability-CEFR alignment models) when choosing or updating textbooks (Chen & Wu, 2024; Yusuf et al., 2024). Lastly, the overall plateau-effect that is evident in textbooks highlights the necessity of teacher professional development in syntactic complexity awareness, which allows teachers to scaffold, monitor, and extend structural development in the learners in ways that would meet the CEFR descriptions of B2 or higher.

Overall, although Intermediate English Textbook 1 can be a reasonable point of departure of the intermediate learners, the given study suggests that the actual development of proficiency, as it is characterized by the CEFR, is impossible without the materials that are actively developed in the complexity of sentence structure and which encourage the learners to apply increasingly challenging syntactic patterns. Pakistani EFL curriculum and materials can better be used to help the learner attain the communicative autonomy and structural complexity that are the characteristic of CEFR-B2 and beyond by closing the gap between the static input and dynamic learner development.

References

- Anggia, H., & Habók, A. (2023). Textual complexity adjustments to the English reading comprehension test for undergraduate EFL students. *Heliyon*, 9(1), e12891.
<https://doi.org/10.1016/j.heliyon.2023.e12891>



- Asif, M., Saeed, A., & Kang, A. (2021). Evaluation of the English Curriculum on the Basis of Linguistics Skills Approved for Higher Secondary Level Public Institutions of Karachi. *Pakistan Social Sciences Review*, 5(2), 588-601.
- Atak, N. (2021). Syntactic complexity in L2 learners' argumentative writing: Developmental stages and the within-genre topic effect. *Journal of Second Language Writing*. <https://doi.org/10.1016/j.jslw.2020.100789>
- Bele, M. (2025). NP Complexity in EFL Argumentative Writing at the CEFR B1, B2 and C1 levels: Corpus-driven Fine-grained Measures. *ELIA: Estudios de Lingüística Inglesa Aplicada*, (25), 217-251.
- British Council. (2022). *English language teaching, learning and assessment in Pakistan: Policies and practices in the school education system*. British Council.
- Chen, X., & Wu, J. (2024). Material design for advanced intermediate learners: Aligning syntactic complexity with proficiency frameworks. *Journal of Language Teaching & Research*, 15(2), 112-126.
- Golparvar, S. E., Casal, J. E., & Abolhasani, H. (2025). The development of syntactic complexity in integrated writing: A focus on fine-grained measures. *Assessing Writing*, 66, 100983.
- Hamid, H. A., Abd. Samad, A., & Noor, N. M. (2025). A systematic review of CEFR-aligned English language textbooks in Malaysia. *Journal of Nusantara Studies*, 10(1), 324–352.
- Hussain, A. (2022). Readability analysis of text in English textbooks of HSSC-II (FBISE). *Linguistic Forum*, 4(3).
- Issa, M., Hussain, A., & Abbas, G. (2022). A quantitative analysis of linguistic complexities: A case study of English textbooks of Federal Board of Intermediate and Secondary Education Islamabad, Pakistan. *Pakistan Languages and Humanities Review*, 6(4), 509–525.
- Khushik, G. A. (2020). Investigating syntactic complexity in EFL learners' writing and the CEFR. *Applied Linguistics*, 41(4), 506–528.
- Kyle, K. (2020). *Measuring syntactic complexity in L2 writing: Using the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC)*. *Language Learning*, 70(3), 665–708.
- Li, C., Wang, X., & Qian, L. (2025). Exploring syntactic complexity and text readability in an ELT textbook series for Chinese English majors. *SAGE Open*. 1-17. <https://doi.org/10.1177/21582440251323619>
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, 45(1), 36–62.
- Mushtaq, H., Bhatti, A. M., & Yasmin, T. (2021). A corpus-based vocabulary analysis of Intermediate Book 1 used in the colleges of Punjab. *Competitive Linguistic Research Journal*, 2(1), 31–57.
- Noor, H., Siddique, A. R., Noor, N., & Latif, T. (2025). Reevaluating textbooks' organization: A corpus-based analysis of content complexity in PCTB English textbooks for intermediate learners. *Journal of Applied Linguistics and TESOL*, 8(2), 1290–1305.
- Saricaoglu, A., & Atak, N. (2022). Syntactic complexity and lexical complexity in argumentative writing: Variation by proficiency. *Novitas-ROYAL (Research on Youth and Language)*, 16(1), 56–73.
- Yusuf, M., Zein, T. T., & Nurlala, N. (2024). Assessing lexical density and its relevance to the CEFR level of reading materials for non-English-major students. *Research Journal in Advanced Humanities*, 5(4). <https://doi.org/10.58256/9ex2es14>
- Zhang, L. J. (2024). EFL students' syntactic complexity development: A latent class growth analysis. *Journal of Second Language Writing*, 55, 101–115. <https://doi.org/10.1016/j.jslw.2024.101115>
- Zhang, X., & Lu, X. (2022). Aligning linguistic complexity with the difficulty of English texts for L2 learners based on CEFR levels. *Studies in Second Language Acquisition*, 44(1), 76–96. <https://doi.org/10.1017/S0272263121000431>