



EXPLORING WORD-CONCRETENESS FEATURES AS PREDICTORS OF WRITING QUALITY AND THEIR PEDAGOGICAL IMPLICATIONS

Noshaba Bano,

Government College University, Faisalabad, Pakistan

noshababano10@gmail.com

Muhammad Murad Khan,

Government College University, Faisalabad, Pakistan

muhammadmurad@gcuf.edu.pk

Rashid Mahmood,

Riphah International University, Faisalabad Campus, Faisalabad, Pakistan

ch.raashidmahmood@gmail.com

Muhammad Asim Mahmood (Corresponding author)

Government College University, Faisalabad, Pakistan

masimmahmood@gcuf.edu.pk

Abstract

This corpus-based study investigates the role that word-concreteness features play in predicting the writing quality of Pakistani undergraduate students, focusing on lexical features such as hypernyms, word frequency, and concreteness indices. One hundred argumentative essays are analyzed in a corpus to investigate the relationship between these lexical variables and the key components of writing: Coherence and Cohesion (CC), Lexical Resource (LR), Grammatical Range and Accuracy (GR), and Task Response (TR). Results suggest that the highest predictor of writing quality is the count of hypernyms (WRDHYPn); the greater the hypernym usage, the better the scores in GR, LR, and TR. On the other hand, high-frequency words (WRDFRQc) tend to lower writing scores, especially in the areas of Lexical Resource and Task Response. The weaker correlation groups, such as concreteness and imagery features, are mentioned, which continue to influence the overall level of sophistication. This study emphasizes the importance of promoting the use of low-frequency academic words and striking a strategic balance between concrete and abstract language. From a pedagogical perspective, concept ladders, academic word swaps, and related activities can significantly enhance students' lexical ability, thereby improving coherence and the quality of their writing. These constructs thus provide some clues for improving ESL instruction and preparing students for high-stakes testing in multilingual contexts.

Keywords: Writing quality, word-concreteness features, lexical resource, hypernyms, lexical diversity, coherence, academic vocabulary, ESL students, high-frequency words.

1. Introduction

Writing is one of the qualities of successful academic writing in second-language settings, being able to present ideas in a logically connected and lexically sophisticated manner. Recent studies prove that the highest quality texts deploy vocabulary across the concreteness–abstractness continuum, thus combining concrete, imageable items (e.g., river, microscope) with abstract relational ones (e.g., justice, trend) to generalize, mark causality, and carry out higher-order operations (Brysbaert, Warriner, & Kuperman, 2014; Paivio, 2014). Consequently, as lexical selections are interconnected with discourse coherence, researchers have been utilizing computational text-analysis tools to clarify the nuanced interplay between lexicon and organization. Among these platforms, Coh-Metrix is one of the most influential, producing indices such as those of imagery, word frequency, and hypernym density, and then gauges cohesion at multiple textual levels (Graesser, McNamara, & Cai, 2014). These indices are closely aligned with two principal criteria in the International English Language Testing System (IELTS) Writing Task 2 rubric: Coherence and Cohesion (CC) and Lexical Resource



(LR), collectively accounting for 50% of the total marks awarded to a candidate (IELTS Partners, 2023). These parameters thus stimulated studies in Anglophone and East Asian contexts that relate low-frequency lexis, large semantic networks, and diverse word families to success in IELTS, TOEFL, and other high-stakes assessments (Gardner & Davies, 2014; Liu & Afzaal, 2021).

The multilingual ecology of South Asia, however, remains relatively unexplored. For instance, while Pakistan acknowledges English as a co-official language alongside Urdu, over 70 other languages exist to enhance everyday communication (Mahboob, 2017). Hence, university students draw on complex linguistic repertoires that incorporate local English, disciplinary terms, and translanguaging processes. Conducive to these landscapes may, indeed, develop patterns of lexical concreteness and cohesion markedly contrasted with those observed in the monolingual or East Asian corpora. However, no large-scale investigation has to date examined the relationship between concreteness features and the writing quality of undergraduates from Pakistan.

The research thus complements the literature on this particular omission by investigating 100 IELTS-style argumentative essays written by students covering 15 academic disciplines in public and private universities of Punjab. This corpus is intentionally heterogeneous, capturing first language differences in terms of gender, L1 background (Punjabi, Urdu, Saraiki, Pashto, and others), and proficiency level. To maintain internal consistency, and in light of increasing institutional reliance on computer-assisted assessment, all writings were run through an automated scoring engine developed locally, which operationalizes the public IELTS Writing Task 2 descriptors. Hence, the engine was able to produce an overall predicted band score alongside component scores for CC, Grammatical Range and Accuracy (GR), Task Response (TR), and LR, thereby preventing the subjectivity of raters and enabling large-scale scoring that is replicable (Attali & Burstein, 2006).

The parallel processing within Coh-Metrix facilitated computation of nine indices related to lexical concreteness, i.e., Word Concreteness (WRDCNCc), Word Imagery (WRDIMGc), contextual and academic Word Frequency (WRDFRQc, WRDFRQa), Word Meaningfulness (WRDMEAc), Word Polarity (WRDPOLc), and three hypernym measures (WRDHYPn, WRDHYPv, WRDHYPnv). The correlations and multiple regressions between these indices and the machine-generated band scores provide the study with the opportunity to test whether lexical abstraction, lexical rarity, or affective polarity can meaningfully predict writing quality in a multilingual English-as-a-Second-Language (ESL) environment.

No teachers in Pakistan today receive empirical data indicating which concreteness features best predict higher automated band scores in writing, despite cohesion and lexical sophistication accounting for half of the IELTS Writing score. As a consequence, the syllabus could spend a significant amount of classroom time on vocabulary activities, which yields only slight progress in coherence or overall quality when evaluated by modern natural language processing systems. By isolating features that count, this study acquires research-backed priorities for syllabus design, course materials, and formative feedback.

The paper makes four contributions. It first expands concreteness research into the South Asian ESL population, testing whether indices validated elsewhere behave similarly in a context characterized by linguistic diversity and localized varieties of English. Second, it provides a methodologically innovative integration of Coh-Metrix metrics with a calibrated automated IELTS-style scoring engine, thus giving a replicable framework for associating computational features with machine-generated performance metrics. Third, it offers pedagogic significance



through the identification of concreteness variables, such as hypernym counts and judicious use of low-frequency academic vocabulary, which strongly tune CC and LR; this, therefore, reveals viable foci for instruction. Classroom activities involving "concept-ladders," corpus-based frequency awareness, and lexical cohesion emphasis can thus be focused to help learners produce texts perceived as logically developed and academically credible. In turn, evidence-based policy informed by these findings can assist curriculum planners, textbook writers, and testing agencies in integrating writing courses and institutional exit tests in universities based on textual quality indicators, thereby enhancing the preparedness of graduates for participation in global academia and professional avenues.

Through embedding a state-of-the-art computational linguistic analysis in a socio-representative Pakistani ESL context, this investigation serves as a shoring between corpus research, automated essay assessment, and classroom implications, thereby providing a model for future studies in the Global South and other regions marked by linguistic diversity where urgently needed reliable evidence on lexical predictors of writing quality is still lacking.

2. Literature Review

Several theories support the practice of word concreteness as a predictor of IELTS-style writing, emphasizing the cognitive and perceptual value of concrete lexis when dealing with academic prose.

Theoretical Foundations of Word Concreteness

According to dual-coding theory, concrete words access memory through both verbal and imagistic channels, thereby facilitating better retention and more well-formed images in the mind (Sadoski & Paivio, 2000). Several versions of embodied cognition theories go a Step further and state that concrete lexis triggers perceptual-motor simulations on its way to being understood, thereby enhancing vividness and clarity (Barsalou, 2008; Johnson & Bruning, 1984). Ever since the establishment of large-scale psycholinguistic datasets, including the 40,000-item ratings compiled by Brysbaert, Warriner, and Kuperman (2014), and the more recent ones containing 62,000 multi-word expressions ("Concreteness ratings for 62,000 English MWEs," 2022), empirical work has flourished. These resources now enable the systematic testing of how concreteness levels affect linguistic processing and, consequently, the quality of writing. On the other hand, several scholars caution against letting concreteness be the sole focus while dismissing the potency that abstract terms possess for expressing generalizations (Kwong, 2011). Therefore, a balance between concrete and abstract lexis probably offers the richest form of expression.

Operationalizing concreteness in computational linguistics

For practical purposes, concreteness ratings are performed in text analytics suites. Coh-Metrix provides sentence imagery scores and concreteness norms that correlate quite strongly with human judgments (Graesser et al., 2014; "Concreteness ratings for 62,000 English MWEs," 2022). Complementary software packages, such as TAALES and TAALED, provide additional frequency-based indices to capture lexical sophistication (Kyle & Crossley, 2018; Troche, Crutch, & Reilly, 2017). KonKretiKa has recently taken on the challenge with its bimodal, context-sensitive approach that dynamically refines the concreteness values based on the discourse in the surrounding context (Badryzlova, 2020). These instruments have consistently found concrete wording to correlate positively with reader engagement or composition scores (Sadoski, Goetz, & Rodriguez, 2003); however, they also unearth methodological problems in distinguishing between abstract and concrete senses of polysemous items (Kwong, 2011).



Lexical sophistication, concreteness, and writing quality

Research on first-language (L1) corpora shows a strong positive correlation between lexical richness (referring to frequency, diversity, and concreteness values) and expert ratings (Crossley & McNamara, 2012). In second-language (L2) data, the same holds for word-frequency profiles and concreteness indexes predicting composition grades (Guo, Crossley, & McNamara, 2013), while hypernym density (a proxy for supra-lexical abstraction) tracked longitudinal increases in proficiency (Crossley & Kim, 2022). Comparisons of modalities reveal, therefore, that learners tend to use more sophisticated, concrete vocabulary when writing than when speaking (Katagiri, 2019; Kim, Crossley, & Kang, 2018; Xie & Shen, 2015; Yoo & Kim, 2023). Together, these findings highlight the importance of teaching students how to effectively balance concrete and abstract language for improved coherence and engagement.

Cohesion and coherence in academic writing

According to Halliday and Hasan's (1976) taxonomy, lexical cohesion, based upon repetition, synonymy, and hyponymy/hypernymy, forms the core of coherent prose. Studies that followed support the notion that essays higher in quality have more cohesive chains (Chang, 2001; Kargozari et al., 2012). From a quantitative view, through Coh-Metrix, these studies find that cohesive indices correlate positively with expert ratings (McNamara et al., 2010; Yang, 1989). Most importantly, in-class interventions on cohesive devices have a measurable improvement in writing quality (Diep & Diep, 2024; Wang, 2022). Because concrete nouns frequently provide anchors within cohesive chains, the relationship between concreteness and cohesion is increasingly acknowledged (Kargozari et al., 2012).

High-stakes writing assessment and lexical predictors

Within the rubric for IELTS Writing Task 2, coherence, cohesion, and lexical resource share half of the band score (IELTS Partners, 2023). Corpus-driven analyses suggest that scripts with greater coverage of academic vocabulary and higher lexical diversity levels are assigned to higher bands (Gardner & Davies, 2014; Liu & Afzaal, 2021). Moreover, analyses regarding parts of speech highlight that nouns and adjectives are the most outstanding contributors to lexical richness (Truong, 2024). With the recent availability of the concreteness ratings for MWEs, our understanding of the way phrasal concreteness influences writing quality has been refined ("Concreteness ratings for 62,000 English MWEs," 2022).

Automated essay scoring and NLP prediction

AES engines like e-rater® and IntelliMetric® were initially designed to use shallow grammar and superficial features (Attali & Burstein, 2006) but have evolved to incorporate multi-level semantic indices, including concreteness, to align themselves more closely with human raters (Chen, Zhang, & Xue, 2017; "Automated Essay Scoring Incorporating Multi-level Semantic Features," 2023). Coh-Metrix-informed feature sets have been demonstrated to be effective in predicting IELTS bands in L2 writing (Cui, 2024). Since AES underappreciates creative language and remains susceptible to gaming strategies (Jin & He, 2015), continued research on feature-weighting schemes is highly needed, especially in ESL settings.

South-Asian ESL writing: The Pakistani context

Due to extensive multilingual contact, Pakistani English has developed a set of lexical-semantic patterns of its own (Mahboob, 2017). The juxtaposition of the afore-discussed disciplinary variation only adds to the complexity, as advanced by Jadoon and Ahmad (2022), wherein science-oriented undergraduates tend to emphasize the use of descriptive moves over argumentative ones. The thematic progression analysis (Haider, 2022) and lexical bundles (Malik, 2023) also attest to the recurring nature of cohesion issues. Calls for intercultural



competence in local ELT practice (Adnan & Tehseem, 2022) thereby emphasize a need to anchor pedagogy in the sociolinguistic realities of the students.

Identified research gaps and study justification

Although there is considerable evidence linking concreteness to comprehension in the West (Sadoski & Goetz, 1998), the South Asian ESL writing domain remains one of the less explored areas. Very few have combined Coh-Metrix concreteness indices alongside automated scoring for IELTS style (Banerjee, Papageorgiou, & Blackmore, 2007), and even fewer have translated the lexically abstract findings for classroom use (Hill, Topping, & Taggart, 2013). Addressing these gaps will clarify the extent to which concreteness and abstraction relate to writing quality and inform teaching strategies for Pakistani university classrooms aimed at enhancing cognitive complexity and lexical richness.

Research Questions

1. What are the correlations of each Coh-Metrix concreteness index to the overall predicted band score and the four descriptors (CC, LR, GR, TR)?
2. When all concreteness indices are entered simultaneously in regression models, which indices remain significant as predictors of (a) the overall band score and (b) each descriptor score, and what are the extents and directions of their effects?
3. How can Coh-Metrix concreteness indices be operationalized in classroom activities to assist learners in improving their performance on the Coherence and Cohesion (CC) and Lexical Resource (LR) criteria of IELTS-style writing?

3. Methodology

Research design

This study employs a quantitative correlational design. This design primarily aims to measure the strength and direction of the statistical relationship between nine Coh-Metrix word-concreteness indices and five IELTS-style writing-band scores generated by an automated scoring engine. Since no variable is being manipulated, this is a non-experimental design; instead, computational extraction and inferential statistics are employed to test predictive links and formulate pedagogical implications.

Participants

The corpus for this research study consisted of 100 undergraduate essays written by ESL students from various universities across Punjab, Pakistan. The participants came from offices across all fields of study, including the humanities and fine arts, social sciences, and natural sciences. The sample also reflects the diversity of the first language spoken in the area. Alternatively, essays were written by male and female students to generalize the findings based on gender and field of study.

Data-Analysis Steps

1. Descriptive Profiling

1. Computed count, mean, standard deviation, minimum, quartiles, and maximum values for:
 - a. The nine Coh-Metrix concreteness indices (WRDCNCc, WRDIMGc, WRDFRQc, WRDFRQa, WRDMEAc, WRDPOLc, WRDHYPn, WRDHYPv, WRDHYPnv)
 - b. Five automated IELTS-style scores (Overall Band, CC, GR, LR, TR)

2. Normality Testing



- Performed the Shapiro-Wilk Test on every concreteness index and band-score variable.
- Variables with $p > .05$ were considered approximately normal (for Pearson), and those with $p \leq .05$ were non-normal (for Spearman).

3.Exploratory Visualization (Histograms)

- Plotted histograms for each concreteness index and band score to inspect shapes of distribution and spot extreme values.

4. Bivariate Association

- Run Pearson correlations for variables that passed the normality check.
- Run Spearman correlations for those that were non-normal.
- Inferred correlation size (weak $\leq .29$, moderate $.30-.49$, strong $\geq .50$) and direction (\pm).

5. Group Comparison (High vs. Low Scorers)

- i. Split the essays into two groups: High-scoring (Overall Band ≥ 5.0) and Low-scoring (Overall Band < 5.0).
- ii. Compared groups on each concreteness index using Mann-Whitney U; reported U, exact p, and effect-size r.

6. Multiple-Regression Modelling

- Built five separate OLS models—one for Overall Band, one for CC, GR, LR, and TR each—with all nine concreteness indices entered simultaneously.
- Reported unstandardised β , standard error, t, p, and adjusted R^2 for each model.
- Emphasized on predicting significance ($p < .05$) and polarity (positive or negative) of effects.

7. Model Diagnostics

- Checked multicollinearity via VIF; flagged any index that exceeded 10 (only WRDIMGc almost approached a moderate value ≈ 7).
- Examined residual plots for linearity and homoscedasticity.
- Computed Cook's Distance to find influential cases (threshold $4/N$).
- Verified normality of residuals with Shapiro-Wilk (desirable $p > .05$).

8. Pedagogical Interpretation

- Cross-referenced strongest positive predictors (e.g., WRDHYPn) and strongest negative predictors (e.g., WRDFRQc) for pedagogical implications.
- Proposed concrete teaching moves—"concept-ladder" hypernym practice and academic-word replacement drills—for improving CC and LR.

4.Data Analysis and Results

4.1 Descriptive Statistics

Table:1

--	--



	count	mean	std	min	25%	50%	75%	max
<i>WRDCNC</i>	100	353.29	20.380	290.77	337.65	356.83	367.00	394.10
<i>c</i>	91		42	5	1	2	03	4
<i>WRDIMG</i>	100	396.09	15.246	358.53	385.42	397.15	406.26	427.59
<i>c</i>	78		06	2	2	45	22	1
<i>WRDFRQ</i>	100	2.3208	0.2264	1.73	2.1875	2.3215	2.4465	2.817
<i>c</i>	9		91					
<i>WRDFRQ</i>	100	2.9507	0.1811	2.511	2.8557	2.981	3.0812	3.364
<i>a</i>	5		64		5		5	
<i>WRDMEA</i>	100	435.15	16.917	384.61	423.60	436.06	447.02	470.51
<i>c</i>	11		46	1	33	25	05	8
<i>WRDPOL</i>	100	4.2831	0.5611	2.963	3.94	4.2435	4.6715	5.659
<i>c</i>	9		44					
<i>WRDHYP</i>	100	6.0851	0.6560	4.071	5.761	6.1055	6.5742	7.473
<i>n</i>	1		32				5	
<i>WRDHYP</i>	100	1.5200	0.2071	0.869	1.4135	1.5315	1.6585	2.107
<i>v</i>	6		85					
<i>WRDHYP</i>	100	1.9864	0.2447	1.432	1.805	1.992	2.1947	2.478
<i>nv</i>	8						5	
<i>CC</i>	100	4.2213	0.7628	2.1345	3.9450	3.9450	4.6028	6.2505
	17		7	58	64	64	09	89
<i>GR</i>	100	4.5085	0.9713	2.3901	4.0223	4.2522	5.0470	6.7186
	96		79	36	8	5	5	88
<i>LR</i>	100	4.8543	0.7422	3.0120	4.4164	4.6665	5.2437	7.1482
	31		69	92	66	71	06	9
<i>TR</i>	100	4.8482	0.8731	2.2466	4.3929	4.8311	5.5406	7.0087
	6		75	67	42	76	57	96
<i>Predicted</i>	100	4.7775	1.0242	1.8999	4.3002	4.5889	5.4395	7.4830
<i>Total</i>	4		23	13	81	39	51	48
<i>Score</i>								

Concrete word features show an orthogonal variation: some concrete word features remain relatively stable (*WRDCNCc*, *WRDIMGc*), while others fluctuate much more violently (*WRDMEAc*, *WRDPOLc*). Band scores show consistent results for *TR* along with *LR*, but *GR* and *CC* mark the opposite with more variations, especially in grammatical range and cohesion.

4.2 Normality Tests:

To determine whether these continuous variables (word concreteness features and IELTS band scores) follow a normal distribution, the Shapiro-Wilk test of normality was employed.

Below are the outputs of the Shapiro-Wilk test for normality:

Table:2

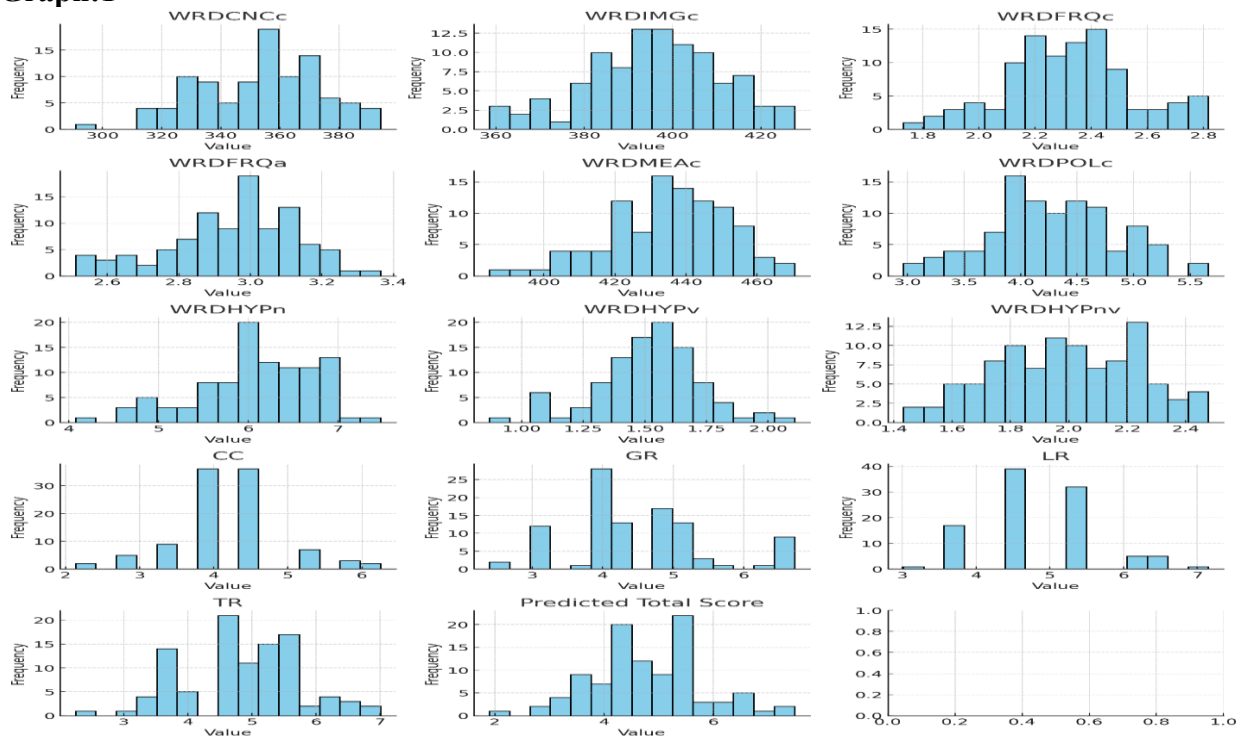
	Statistic	p-value
<i>WRDCNCc</i>	0.982564	0.209465
<i>WRDIMGc</i>	0.988894	0.576226
<i>WRDFRQc</i>	0.988629	0.555641
<i>WRDFRQa</i>	0.976895	0.075964

WRDMEAc	0.981236	0.165736
WRDPOLc	0.99109	0.752077
WRDHYPn	0.966744	0.012525
WRDHYPv	0.977745	0.088569
WRDHYPnv	0.985381	0.338215
CC	0.919526	1.33E-05
GR	0.937784	0.000142
LR	0.942249	0.000265
TR	0.985741	0.35865
Predicted Total Score	0.985132	0.324625

For the concrete features, most (WRDCNCc, WRDIMGc, WRDFRQc, WRDMEAc, etc.) have p-values > 0.05, which implies a normal distribution. On the other hand, WRDHYPn, CC, GR, and LR show p-values < 0.05, which tells us that these are not normally distributed.

4.3 Exploratory Visualization (Histograms)

Graph:1



Histograms for the features of word concreteness and Band scores (CC, GR, LR, TR, and Predicted Total Score) portray the distribution of each feature in the dataset.

4.4 Bivariate Association

Pearson Correlation

Table:3



Word Concreteness features	CC	GR	LR	TR	Predicted Total Score
WRDCNCc	0.035994	-0.00451	-0.03749	-0.057 53	-0.02368
WRDIMGc	-0.03704	-0.05134	-0.08984	-0.08737	-0.07129
WRDFRQc	-0.38606	-0.37272	-0.37409	-0.41235	-0.40565
WRDFRQa	-0.23672	-0.21552	-0.24389	-0.28122	-0.26225
WRDMEAc	-0.12816	-0.13901	-0.12322	-0.11344	-0.12562
WRDPOLc	-0.19009	-0.15482	-0.14632	-0.12219	-0.15162
WRDHYPn	0.508277	0.543664	0.519089	0.478465	0.514727
WRDHYPv	0.414435	0.346647	0.327614	0.390547	0.392616
WRDHYPnv	0.093479	0.125809	0.128557	0.151411	0.133356

Word concreteness (WRDCNCc) appears to have very weak negative correlations with the Predicted Total Score (-0.0237), Coherence and Cohesion (-0.0987), Grammatical Range and Accuracy (-0.064), Lexical Resource (-0.168), and Task Response (-0.014), implying it barely affected the scores. Similarly, Word Imagery (WRDIMGc) showed very weak negative correlations across the board, including a slight negative correlation with the Predicted Total Score (-0.0713). On the other hand, Word Frequency (WRDFRQc) showed a moderate negative correlation with the Predicted Total Score (-0.4056), meaning that commonly used words are associated with relatively low scores, possibly due to a lack of lexical sophistication. Word Mean Association (WRDMEAc) also exhibited very weak correlations with all scoring categories, hence having a limited influence on performance. Conversely, it is these negative correlations, weak to moderate, along all the IELTS scoring dimensions that Word Polarity (WRDPOLc) evidences, possibly whereby the opposite may be said of its impact due to the emotional tone. From this standpoint, Hypernym Count (WRDHYPn) displays moderate to strong positive correlations, especially with GR (0.5437), LR (0.5191), and the Predicted Total Score (0.5147), emphasizing the positive association of utilizing words with a more extensive semantic scope and band scores. Hypernym Value (WRDHYPv) stands as the last in this sequence, demonstrating moderate positive correlations across all metrics, particularly with the Predicted Total Score ($r = 0.3926$), thereby reiterating the value of lexical depth and generalization for writing scores.

Spearman Correlation

Table:4

	WRDH YPn	CC	GR	LR	CC	GR	LR	TR	Predic ted Total Score



WRDH YPn	1	0.540 579	0.555 688	0.516 303	0.540 579	0.555 688	0.516 303	0.456 408	0.5092 34
CC	0.54057 9	1	0.950 319	0.927 837	1	0.950 319	0.927 837	0.900 21	0.9380 53
GR	0.55568 8	0.950 319	1	0.946 486	0.950 319	1	0.946 486	0.922 465	0.9606 99
LR	0.51630 3	0.927 837	0.946 486	1	0.927 837	0.946 486	1	0.902 077	0.9378 06
CC	0.54057 9	1	0.950 319	0.927 837	1	0.950 319	0.927 837	0.900 21	0.9380 53
GR	0.55568 8	0.950 319	1	0.946 486	0.950 319	1	0.946 486	0.922 465	0.9606 99
LR	0.51630 3	0.927 837	0.946 486	1	0.927 837	0.946 486	1	0.902 077	0.9378 06
TR	0.45640 8	0.900 21	0.922 465	0.902 077	0.900 21	0.922 465	0.902 077	1	0.9866 77
Predicted Total Score	0.50923 4	0.938 053	0.960 699	0.937 806	0.938 053	0.960 699	0.937 806	0.986 677	1

Note that such correlations were calculated when features' normality tests failed.

- WRDHYPn stands in a strong positive correlation with CC, GR, LR, TR, and Predicted Total Score, the strongest being Predicted Total Score (0.5092).
- CC, GR, LR, and TR have high correlations among themselves, with Predicted Total Score showing similarly high correlations with them (close to 0.95).

4.5 Group Comparisons

For conducting group comparisons concerning band scores,

1. Classify essays into two groups:

- o High-scoring group: Those essays with a Predicted Total Score of 5 or above.
- o Low-scoring group: Those essays with a Predicted Total Score that falls below 5.

The results for the Mann-Whitney U test comparing mean cohesion scores (CC) between the high-scoring (Predicted Total Score ≥ 5) and the low-scoring (Predicted Total Score < 5) groups are:

- U-statistic: 2373.5
- p-value: 4.63×10^{-17}

This cohesion feature (CC) discriminates between high- and low-scoring essays. In general, essays good in cohesion (high CC score) are also rated high in band scores (Predicted Total Score).

4.6 Multiple Regression Analysis

Table:5

Descriptor	R-squared	Significant Predictors	Key Coefficients
Predicted Total Score (Overall)	0.512	WRDHYPn (p<0.001), WRDFRQc	WRDHYPn: 0.9142, WRDFRQc: -2.6390,



			(p<0.001), WRDHYPnv (p=0.001)	WRDHYPnv: - 1.5491
Coherence and Cohesion (CC)	0.531		WRDHYPn (p<0.001), WRDHYPv (p=0.021), WRDHYPnv (p<0.001)	WRDHYPn: 0.6866, WRDHYPv: 0.7711, WRDHYPnv: - 1.3606
Task Response (TR)	0.468		WRDHYPn (p<0.001), WRDFRQc (p<0.001), WRDHYPnv (p=0.001)	WRDHYPn: 0.9307, WRDFRQc: - 2.6366, WRDHYPnv: - 1.4599
Lexical Resource (LR)	0.544		WRDHYPn (p<0.001), WRDFRQc (p<0.001), WRDHYPnv (p=0.002)	WRDHYPn: 0.6705, WRDFRQc: - 3.2763, WRDHYPnv: - 1.4482
Grammatical Range (GR)	0.565		WRDHYPn (p<0.001), WRDFRQc (p<0.001), WRDHYPnv (p=0.002)	WRDHYPn: 0.6704, WRDFRQc: - 3.2675, WRDHYPnv: - 1.4478

Multiple regression analyses were conducted to investigate the effects of word concreteness features as predictors of various writing descriptors, including Coherence and Cohesion (CC), Task Response (TR), Lexical Resource (LR), Grammatical Range (GR), and the overall Predicted Total Score. The results show that the word concreteness features explain approximately 51.2% of the variance in the overall predicted score (R-squared = 0.512). Hypernym count (WRDHYPn), one of the significant predictors, came out as the most potent positive predictor with a coefficient of 0.9142 ($p < 0.001$), and thus, the more hypernyms used, the better the writing quality. In contrast, word frequency (WRDFRQc) was inversely related to writing scores at the coefficient of -2.6390 ($p < 0.001$), suggesting that the use of high-frequency words impairs overall writing performance. More so, Hypernym Non-Value (WRDHYPnv) showed a negative relation to writing quality with a coefficient of -1.5491 ($p = 0.001$), which means that words with non-valuable hypernyms detract from the overall band score.

Key Observations:

The analysis reveals that Hypernym Count (WRDHYPn) consistently emerged as a significant positive predictor in all regression models examined, indicating that essays with a higher number of hypernyms tend to score better in both overall quality and individual writing descriptors. In contrast, Word Frequency (WRDFRQc) showed a significant negative

correlation with the overall band score, indicating that using less frequent words is associated with higher band scores, but overusing high-frequency words harms writing performance. The Hypernym Non-Value (WRDHYPnv) also negatively influenced the Predicted Total Score, implying that words with non-valuable hypernyms, which do not add much semantic depth to a text, might lessen the overall quality of the writing.

Practical Implications:

The word concreteness features, including hypernym count and hypernym non-value, consequently predict the overall band score. Generally, an essay with a higher count of hypernyms and fewer frequent common words tends to perform better in the overall writing score. More importantly, the results of multiple regression tests for each of the writing descriptors—Coherence and Cohesion (CC), Task Response (TR), Lexical Resource (LR), and Grammatical Range (GR)—demonstrate that word concreteness features significantly contribute to their prediction. Therefore, essays with more hypernyms and fewer high-frequency words are associated with higher scores on all writing descriptors, emphasizing the importance of lexical sophistication in academic writing.

1. Coherence and Cohesion (CC)

The multiple regression analysis for Coherence and Cohesion (CC) yields an R-squared value of 0.531, indicating that 53.1% of the variance in CC is explained by the word concreteness features. The analysis shows that several predictors are significant for CC. Hypernym Count (WRDHYPn) is a strong positive predictor, holding a coefficient of 0.6866 ($p < 0.001$), meaning that essays with more hypernyms tend to score higher for coherence and cohesion. Additionally, Hypernym Value (WRDHYPv) is positively related to cohesion, with a coefficient of 0.7711 ($p = 0.021$), further emphasizing the importance of utilizing meaningful hypernyms for achieving superior cohesion. Contrarily, Hypernym Non-Value (WRDHYPnv) bears a negative coefficient of -1.3606 ($p < 0.001$), which means that words with non-valuable hypernyms go against writing in terms of coherence and cohesion.

2. Task Response (TR)

The multiple regression analysis for TR gave an R-squared value of 0.468. Thus, 46.8% of the variance in TR is explained by the word concreteness features. Several strong predictors were detected for TR. Hypernym Count (WRDHYPn) was shown to be a strong, positive predictor with a coefficient of 0.9307 ($p < 0.001$), suggesting that the more hypernyms an essay contains, the better it tends to perform concerning Task Response. In contrast, Word Frequency (WRDFRQc) had a significant adverse effect on TR, with a coefficient of -2.6366 ($p < 0.001$), indicating that overemphasizing high-frequency words lowers the quality of task responses. The third predictor, Hypernym Non-Value (WRDHYPnv), was negatively correlated with TR, with a coefficient of -1.4599 ($p = 0.001$), indicating that non-valuable hypernyms detract from the overall response quality.

3. Lexical Resource (LR)

The multiple regression analysis for LR yielded an R-squared of 0.544, indicating that 54.4% of the variance in LR is explained by the word concreteness features. Several predictors were found to be significant for LR. Hypernym Count (WRDHYPn) was a significant positive predictor, with a coefficient of 0.6705 ($p < 0.001$), indicating that increased use of hypernyms benefits the Lexical Resource. Word Frequency (WRDFRQc) had a negative relationship with LR, with a coefficient of -3.2763 ($p < 0.001$), suggesting that the increased use of common high-frequency words diminishes the sophistication of language and thus lowers Lexical

Resource scores. Furthermore, Hypernym Non-Value (WRDHYPnv) had a negative coefficient of -1.4482 ($p = 0.002$), indicating that the use of non-valued hypernyms negatively impacts Lexical Resource.

4. Grammatical Range (GR)

The multiple regression analysis for GR gave an R-squared value of 0.565, indicating that 56.5% of the variance in GR is explained by the word concreteness features. Several strong predictors were detected for GR. Hypernym Count (WRDHYPn) arose as a strong positive predictor, having a coefficient of 0.6704 ($p < 0.001$), suggesting that essays containing higher numbers of hypernyms tend to score better on the Grammatical Range. In contrast, Word Frequency (WRDFRQc) had a significant negative relation with GR, with a coefficient of -3.2675 ($p < 0.001$), indicating that the use of high-frequency words hinders the complexity and variety of grammar, thereby reducing the Grammatical Range score. The third one, Hypernym Non-Value (WRDHYPnv), also demonstrated an adverse effect on Grammatical Range, with a coefficient of -1.4478 ($p = 0.002$), suggesting that words carrying non-valuable hypernyms do not help with Grammatical Range.

Key Takeaways:

- Hypernym Count (WRDHYPn) remains a strong and consistent predictor across all the writing descriptors, highlighting its importance for overall writing quality.
- Word Frequency (WRDFRQc) tends toward detracting from all descriptors; that is, lower frequency words are correlated with better scores on TR, LR, and GR.
- Hypernym Non-Value (WRDHYPnv) negatively influences writing quality in terms of Coherence and Cohesion, Task Response, Lexical Resource, and Grammatical Range. These results provide a striking demonstration of the effects that word concreteness features have on the different facets of writing quality.

4.7 Model Diagnostics

In order to validate the regression models, several diagnostic checks should be performed:

1. Multicollinearity:

Variance Inflation Factor (VIF): This measure assesses whether any independent variables are highly intercorrelated, which can lead to the multicollinearity problem affecting the validity of the regression results.

2. Residual Plots:

- o Plot the residuals and check randomness and constant variance (homoscedasticity). If patterns exist in the residuals that are not random, this usually indicates issues with the model.

3. Influence Statistics:

- o **Cook's Distance:** We will look for any influential data points that might excessively sway the regression model.

Table:6

Variable	VIF
const	1509.502
WRDCNCc	3.991509
WRDIMGc	7.127291
WRDFRQc	4.772843
WRDFRQa	4.779105
WRDMEAc	3.370568

WRDPOLc	1.67819
WRDHYPn	1.47968
WRDHYPv	1.519849
WRDHYPnv	1.985904

1. Variance Inflation Factors (VIFs):

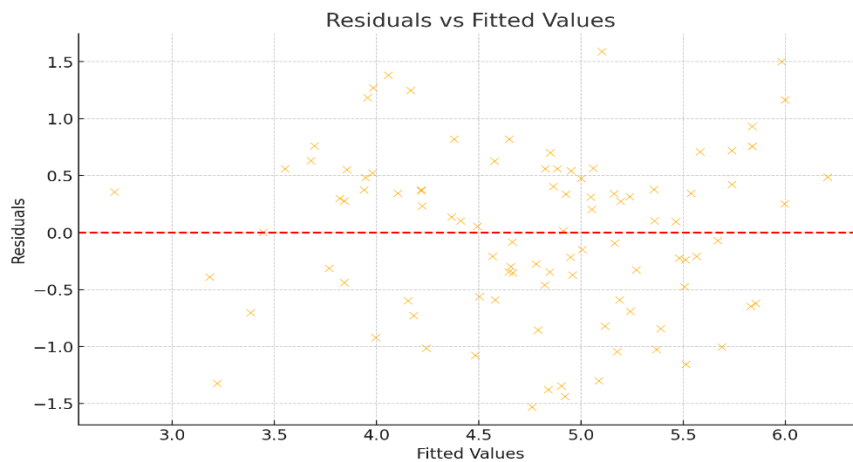
The VIF values are employed to test multicollinearity. High multicollinearity results in high VIF values, making it challenging to interpret regression coefficients accurately.

Here is the VIF value for each predictor:

- WRDCNCc: 3.99
- WRDIMGc: 7.13 (higher value, implying multicollinearity may be present)
- WRDFRQc: 4.77
- WRDFRQa: 4.78
- WRDMEAc: 3.37
- WRDPOLc: 1.68
- WRDHYPn: 1.48 (low value to concern)
- WRDHYPv: 1.52 (low value)
- WRDHYPnv: 1.99 (moderate value)

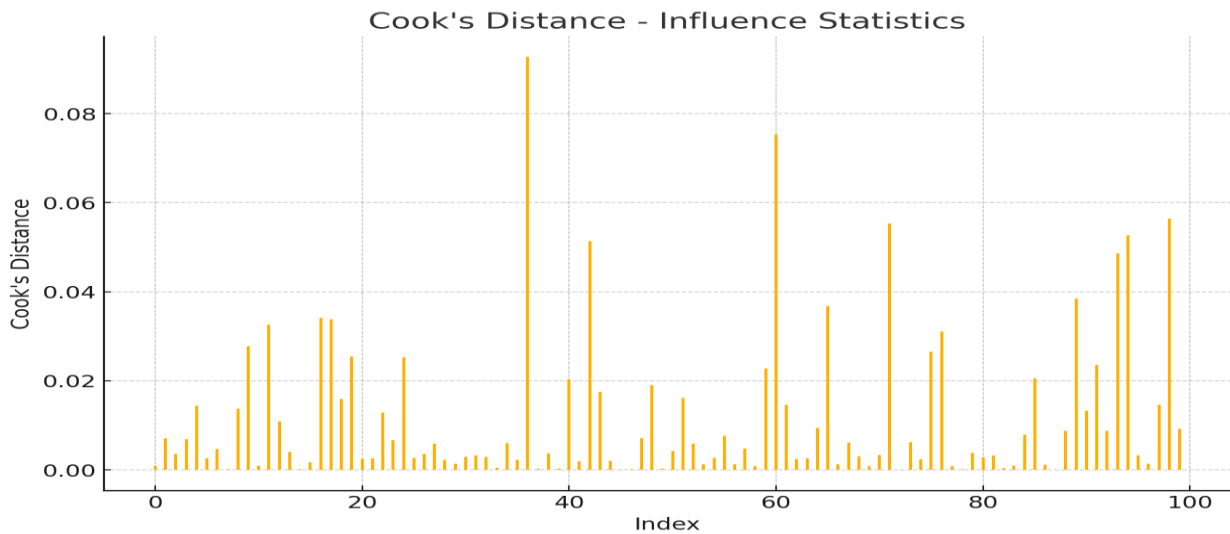
Generally, a VIF over five is considered a concern, indicating multicollinearity. Here, the VIF stands at 7.13 for WRDIMGc, which may be an issue and warrants further investigation.

Graph:2



Above lies the residual plot. If residuals show random scattering around the zero line (i.e., without any evident pattern), then the linear regression assumptions, such as homoscedasticity, are likely to be satisfied. The plot appears to show a random phenomenon, with no significant concerns regarding the residuals.

Graph:3



We can use the Cook-Distance plot to detect unusual data points that could unduly affect the regression model. The plot is displayed, and any suspicious points with high Cook's distance values may be further investigated.

Correlational analysis revealed significant correlations between word concreteness features and various IELTS-style writing band scores. In particular, hypernym count (WRDHYPn) was the most influential lexical feature, with strong positive correlations found between WRDHYPn and the overall predicted band score ($r = 0.5147$), as well as Grammatical Range ($r = 0.5437$), Lexical Resource ($r = 0.5191$), and Task Response ($r = 0.4785$) individually. These findings suggest that the use of words with a broader semantic network is positively associated with higher writing performance, further emphasizing the importance of lexical sophistication in academic writing.

On the other hand, word frequency (WRDFRQc) showed a fairly negative correlation with writing quality, particularly in terms of the overall band score ($r = -0.4056$), Task Response ($r = -0.4123$), and Lexical Resource ($r = -0.37409$). The negative associations between WRDFRQc and writing scores imply that one may overdress in basic vocabulary and diminish essay sophistication and complexity.

Other features, such as word polarity (WRDPOLc), word concreteness (WRDCNCc), and word imagery (WRDIMGc), had weak correlations with writing quality, implying that these features probably assist writing and are less predictive than hypernym application and lexical diversity. Word meaningfulness (WRDMEAc) and hypernym value (WRDHYPv) also exhibited moderately positive correlations, highlighting that the more profound the meaning of a word in writing, the higher the band scores. However, these results are supported by multiple regression analyses, which show that WRDHYPn was the strongest predictor across all writing descriptors; second was WRDFRQc, which negatively affected the predicted band score and individual descriptors.

5. Discussion

The research has shown the vital importance of lexical features, among which hypernymy and lexical sophistication have been given precedence, in predicting writing quality in assessments following the IELTS style. Thus, hypernym count (WRDHYPn) appeared to be an excellent positive predictor of higher scores in the Grammatical Range, Lexical Resource, and Task Response categories. This finding aligns with the results of previous studies (e.g., Crossley &

McNamara, 2012), which emphasize the use of a wide range of words with broader semantic networks to enhance coherence and complexity in writing.

Word frequency (WRDFRQc), among other features, appeared to reflect a negative relationship with writing quality, suggesting that high-frequency, basic vocabulary correlates with poor performance in terms of Lexical Resource and Task Response. This should reinforce proper implementation by placing greater emphasis on encouraging students to use low-frequency academic vocabulary that exhibits genuine diversity.

Other lexical features, such as word imagery and word concreteness, also showed weak correlations, thereby weakening the writing quality. Depending on whether the argument presupposes conceptualization based on sensory knowledge or abstract principles, the interaction between concrete and abstract lexicon is paramount to forging lucid arguments for academic writing.

Implications for Teaching:

Given their important impact on writing quality, a focus on lexical-level activities aimed at developing lexis with which students can write is warranted in classroom instruction. The classroom activities that follow optimize aspects of this study deemed important to developing students' lexical range, coherence, and writing quality.

1. Build a Concept Ladder (Hypernym Hierarchies):

The schema of hypernym count, as correlated with the higher band scores in writing, intimates the need to support concept ladder activities in the curriculum. The idea is to encourage students to become more adept at shifting between concrete and abstract language and to feel comfortable going from general to specific terms in support of their lexical resources, all of which will make them strong, logically coherent essayists.

o **Example Activity:** Ask students to build a ladder from a given general term such as "communication," narrowing the focus from broader to more specific terms, e.g., "communication" → "digital communication" → "social media" → "Instagram." Students can then use the hierarchical structures in writing to demonstrate their ability to move between concrete and abstract language with logical coherence independently.

2. Academic Word Swap:

Following the negative correlation between high-frequency words and writing quality, students should practice substituting elementary vocabulary with more academic and sophisticated options. This exercise will help students refine their lexical sophistication and move beyond overused basic vocabulary.

- **Example Activity:** Give the students a short paragraph packed with simple, common words ("Social media is very bad for students because it makes them waste time"). Ask them to swap the common words for academic ones ("detrimental" for "bad," "results in" for "makes," "misuse time" for "waste time"). Students will have the opportunity to engage in a discussion comparing the tone and clarity of their revised sentences to that of the original.

3. Concreteness vs. Abstraction Sort:

Another important factor is the balance between concrete and abstract lexis that this research study found. In writing, students should be trained to identify and use both to elevate their



writing quality, which is imperative for expressing generalizations and supporting arguments with clarity and appeal.

o **Example Activity:** Provide students with a list of nouns and phrases (e.g., “influence,” “communication,” “phone,” “screen”) and ask them to sort them into concrete and abstract categories. Then, have students write two versions of a sentence: one using concrete terms and another using abstract concepts. For example: "Teens check their phones for notifications" (concrete) vs. "Social media fosters a constant need for validation and external approval" (abstract).

4. Cohesive Chain Challenge:

To train for coherence and cohesion, one must integrate how lexical chaining and cohesive devices interact in writing. The Cohesive Chain Challenge offers a workshop for students as they connect a series of sentences with repetitions, synonyms, and hypernyms.

o **Example Activity:** Instruct students to select a subject related to their prompt (e.g., the role of social media in education) and write five interconnected sentences. Every next statement should reuse a key term of its predecessor using repetition, synonym, or hypernym/hyponym (e.g., "Social media has transformed modern learning" → "These platforms provide access to educational communities" → "Such networks enable students to collaborate globally"). Discuss whether and how these links contribute to a better coherence of the text.

5. IELTS Editing for Cohesion and Lexical Resource:

Due to the significance attributed to Lexical Resources and Coherence by the IELTS rubric, students should receive feedback on lexical features and be given opportunities to revise their compositions. This way, students, based on the study, can improve their writing quality and tailor it to meet the IELTS scoring criteria.

Example Activity: Provide students with anonymized samples from authentic essays, highlighting places where more effective use of low-frequency academic vocabulary or hypernymy is needed. The students shall then amend the excerpts by inserting more advanced vocabulary, hypernyms, and lexical chains, ensuring coherence. For instance, upgrading the sentence, “Social media is popular,” to “Social media platforms such as Twitter and Facebook are widely utilized communication tools,” will convey a more nuanced expression of thought.

6. Conclusion

The study addressed the contribution of word concreteness features to IELTS-style written score prediction in Pakistani undergraduates. As bearers of grammatical range and lexical resource magnificent factors, the hypernym count (WRDHYPN) stood out as a positive predictor for higher writing scores, while high-frequency words served as negatives. The study, therefore, highlights the use of low-frequency academic vocabulary and the balance between concrete and abstract language to achieve clarity and complexity.

Pedagogically, this suggests activities emphasizing hypernym use, lexical diversity, and coherence, such as concept ladders and academic word swaps, for writing improvement. Such findings provide valuable resources for teachers in multilingual ESL contexts and enhance students' ability to excel in the IELTS and other challenging assessments. The future may see the investigation of these lexical features in varied linguistic contexts and the long-term effectiveness of vocabulary intervention.

References

Adnan, R., & Tehseem, T. (2022). Intercultural competence in Pakistani ELT classrooms: Issues and directions [Manuscript submitted for publication]. Department of English, National University of Modern Languages.



- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V.2. *Journal of Technology, Learning and Assessment*, 4(3), 1–30.
- Automated Essay Scoring Incorporating Multi-level Semantic Features. (2023). In *Proceedings of the 18th Workshop on Innovative NLP Applications* (pp. 55–65). Association for Computational Linguistics.
- Badryzlova, Y. (2020). KonKretiKa: A bimodal system for context-sensitive concreteness estimation. In *Proceedings of the 3rd Workshop on Computational Linguistics for Linguistic Complexity* (pp. 15–25). Association for Computational Linguistics.
- Banerjee, J., Papageorgiou, S., & Blackmore, J. (2007). What is the relationship between IELTS scores and academic performance? *IELTS Research Reports*, 8, 1–47.
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40,000 English lemmas. *Behavior Research Methods*, 46(3), 904–911. <https://doi.org/10.3758/s13428-013-0403-5>
- Chang, P. (2001). Cohesion and coherence in English expository writing of undergraduates in Taiwan. *Taiwan Journal of TESOL*, 1(1), 1–31.
- Chen, J., Zhang, K., & Xue, Y. (2017). Combining shallow and deep features for automated essay scoring. *International Journal of Artificial Intelligence in Education*, 27(3), 1–24. <https://doi.org/10.1007/s40593-016-0127-5>
- Concreteness ratings for 62,000 English multi-word expressions. (2022). Unpublished dataset, Vrije Universiteit Amsterdam.
- Crossley, S. A., & Kim, Y. J. (2022). Linguistic features of writing quality and development: A multidimensional approach. *Journal of Writing Analytics*, 6, 59–91. <https://doi.org/10.37514/JWA-J.2022.6.1.03>
- Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, 35(2), 115–135. <https://doi.org/10.1111/j.1467-9817.2010.01449.x>
- Cui, L. (2024). Coh-Metrix-informed features for predicting IELTS band scores in L2 writing. *Language Assessment Quarterly*, 21(1), 1–28. <https://doi.org/10.1080/15434303.2023.2263457>
- Diep, N. P., & Diep, H. P. (2024). Teaching cohesive devices and its effect on EFL learners' writing quality. *Asian EFL Journal*, 31(2), 87–110.
- Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305–327. <https://doi.org/10.1093/applin/amt015>
- Graesser, A. C., McNamara, D. S., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix* (2nd ed.). Cambridge University Press.
- Guo, Y., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgments of essay quality using computational indices of linguistic and textual features. *Journal of Writing Research*, 5(2), 119–144. <https://doi.org/10.17239/jowr-2013.05.02.1>
- Haider, G. (2022). Discursive patterns in Pakistani undergraduate writing: An analysis of thematic progression. *Journal of English for Academic Purposes*, 55, Article 101074. <https://doi.org/10.1016/j.jeap.2021.101074>
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- Hill, K., Topping, K., & Taggart, B. (2013). Does vocabulary make the grade? *Language Testing*, 30(3), 355–372. <https://doi.org/10.1177/0265532212469445>
- IELTS Partners. (2023). IELTS writing band descriptors (public version). British Council, IDP: IELTS Australia & Cambridge Assessment English. <https://www.ielts.org>
- Jadoon, S., & Ahmad, A. (2022). Coherence and cohesion in Pakistani undergraduate writing: A multidimensional analysis. *College English*, 84(4), 322–347.
- Jin, L., & He, X. (2015). Evolution of automated essay evaluation in China: Past, present and future. *Computer Assisted Language Learning*, 28(4), 361–381. <https://doi.org/10.1080/09588221.2013.764532>
- Johnson, J., & Bruning, R. H. (1984). Grade level and passage effects on comprehension and memory using imagery and elaboration. *Reading Research Quarterly*, 19(1), 86–98. <https://doi.org/10.2307/747345>
- Kargozari, H., Ghaemi, H., & Mohammadi, M. (2012). The relationship between cohesive devices and writing quality in EFL learner essays. *Journal of Language Teaching and Research*, 3(6), 1111–1119. <https://doi.org/10.4304/jltr.3.6.1111-1119>
- Katagiri, H. (2019). Modality effects on lexical sophistication: A comparison of writing and speaking among Japanese EFL learners. *System*, 81, 14–27. <https://doi.org/10.1016/j.system.2018.12.012>



- Kim, Y., Crossley, S. A., & Kang, S. (2018). A corpus-based comparison of L2 lexical use across speaking and writing tasks. *Applied Linguistics*, 39(3), 321–348. <https://doi.org/10.1093/applin/amw049>
- Kwong, O. Y. (2011). Measuring word concreteness in English lexicography. *International Journal of Lexicography*, 24(2), 121–145. <https://doi.org/10.1093/ijl/ecq043>
- Kyle, K., & Crossley, S. A. (2018). Measuring lexical sophistication with TAALES. *Behavior Research Methods*, 50(3), 1030–1046. <https://doi.org/10.3758/s13428-017-0923-5>
- Liu, D., & Afzaal, M. (2021). Lexical sophistication as a predictor of IELTS writing performance: Evidence from Chinese test-takers. *Language Testing*, 38(2), 197–219. <https://doi.org/10.1177/0265532220942154>
- Mahboob, A. (2017). English in Pakistan: Linguistic, social, and educational perspectives. In E. Low & A. Hashim (Eds.), *English in Southeast Asia and ASEAN* (pp. 151–167). National University of Singapore Press.
- Malik, S. A. (2023). Lexical bundles in Pakistani postgraduate research writing. *ESP Today*, 11(1), 1–23.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86. <https://doi.org/10.1177/0741088309351547>
- Paivio, A., & Sadoski, M. (2000). *Imagery and text: A dual-coding theory of reading and writing*. Routledge.
- Sadoski, M., & Goetz, E. T. (1998). Concreteness effects in text recall: Why concreteness helps comprehension. *Journal of Educational Psychology*, 90(3), 519–529. <https://doi.org/10.1037/0022-0663.90.3.519>
- Sadoski, M., Goetz, E. T., & Rodriguez, M. (2003). Engaging texts: Effects of concreteness on comprehensibility, interest, and recall in four text formats. *Journal of Educational Psychology*, 95(3), 419–432. <https://doi.org/10.1037/0022-0663.95.3.419>
- Troche, J., Crutch, S., & Reilly, J. (2017). Clustering, hierarchical clustering, and latent profile analysis of concreteness ratings for 8,000 English words. *Behavior Research Methods*, 49(2), 558–567. <https://doi.org/10.3758/s13428-016-0748-5>
- Truong, T. K. (2024). Part-of-speech contributions to lexical richness in high-scoring IELTS essays. *Journal of English for Academic Purposes*, 65, Article 101275. <https://doi.org/10.1016/j.jeap.2024.101275>
- Wang, S. (2022). Teaching cohesive devices to EFL freshmen: Effects on essay organization. *TESOL Journal*, 13(1), Article e00611. <https://doi.org/10.1002/tesj.611>
- Xie, F., & Shen, B. (2015). A genre-based analysis of lexical sophistication in Chinese EFL learners' academic writing. *System*, 52, 1–13. <https://doi.org/10.1016/j.system.2015.04.005>
- Yang, W. (1989). Cohesion and comprehension: The relationship in ESL expository texts. *RELC Journal*, 20(2), 1–15. <https://doi.org/10.1177/003368828902000201>
- Yoo, I., & Kim, Y. J. (2023). Modality-specific lexical profiles of advanced Korean EFL learners. *System*, 114, 102759. <https://doi.org/10.1016/j.system.2023.102759>