

---

**A Corpus-Based Analysis of Narrativity Features and Their Impact on Writing Performance among Pakistani ESL Students**

**Noshaba Bano<sup>1</sup>, Fatima Zia<sup>2</sup>, Muhammad Asim Mehmood<sup>3</sup>, Luqman Manzoor<sup>4</sup>**

1 Department of Applied Linguistics, Government College University Faisalabad, Pakistan.

Email: noshababano10@gmail.com

2 MPhil Scholar, Government College University Faisalabad. fatimazia211@gmail.com

3 Professor, Department of Applied Linguistics, Government College University Faisalabad, Pakistan. Email: masimmahmood@gcuf.edu.pk

4 Department of Applied Linguistics, Government College University Faisalabad, Pakistan.

Email: luqmanmanzoor875@gmail.com

**DOI: <https://doi.org/10.70670/sra.v4i2.2349>**

**Abstract**

This corpus-based study examines the influence of discourse-level narrativity features specifically verb cohesion (PCVERBz), narrative cohesion (PCNARz), and causal reasoning (SMCAUSv, SMCAUSlsa) on the academic writing performance of Pakistani English as a Second Language (ESL) undergraduate students. A localized corpus of 100 argumentative essays was collected from universities across Punjab, Pakistan, and 26 multi-level linguistic indices were extracted using the computational tool Coh-Metrix. The essays were evaluated using local automated scoring software calibrated to the official IELTS Task 2 band descriptors across four core domains: coherence and cohesion (CC), lexical resource (LR), grammatical range and accuracy (GR), and task response (TR). Due to severe violations of normality among the data distributions, Spearman's rank correlation was applied to establish bivariate associations. Concurrently, ridge regression analysis was executed to model the multi-feature predictive relationships while controlling for extreme multicollinearity and overfitting among the cohesive metrics. Bivariate and regression analyses demonstrate that an elevated frequency of repetitive verbs (PCVERBz) consistently predicts lower writing quality across all evaluation descriptors. Conversely, narrative cohesion (PCNARz) and forward causal reasoning via latent semantic analysis (SMCAUSlsa) exhibit strong positive effects on text coherence (CC) and task fulfillment (TR). These findings suggest that English for Academic Purposes (EAP) pedagogy must transition from traditional structural grammar tracking to explicit discourse-level interventions. Pedagogical applications emphasize the implementation of targeted vocabulary substitution drills, cause-and-effect cognitive mapping, and explicit exercises guiding the genre shift from sequential narrative structures to analytical expository prose.

**Keywords:** narrativity features; Coh-Metrix; Pakistani ESL writing; automated essay assessment; discourse-level cohesion; ridge regression

**1. Introduction**

Writing proficiency is regarded as the primary skill for ESL learners, who require it to access the academic avenues available to them in the globalized environment where the English

language is the dominant medium in higher education and international examinations. Coherent writing enables students to convey highly complex ideas, communicate well in general coursework, undertake standardized testing, or simply engage in academic pursuits on a global platform. The writing test is highly instrumental for Pakistani ESL students, as assessments such as IELTS serve as a gateway to both academic and professional progress. IELTS assesses writing output according to its primary descriptors, namely Coherence and Cohesion (CC), Lexical Resource (LR), Grammar (GR), and Task Response (TR) (Lynch, 2001; Weigle, 2002).

While much research is devoted to the influence of lexical diversity, syntactic complexity, and grammatical correctness on writing performance (Crossley & McNamara, 2011), fewer studies address narrativity features such as verb cohesion, narrative structure, and causal reasoning that are responsible for the logical flow and coherence of discourse in writing. Academic writing is crucial for demonstrating the ability to write structured texts that sustain an argument. Examining how narrativity can influence effective writing will bring new perspectives to the discourse-level skills required for high-stakes writing tasks.

Addressing this issue, the present study uses Coh-Metrix, a computational tool that analyses texts at multiple linguistic levels, including narrativity and cohesion (Graesser, McNamara, & Kulikowich, 2004). The research further investigates how Pakistani ESL students utilise narrativity features in their writing and the correlation between these features and performance, as measured by an IELTS-style band descriptor. One hundred undergraduate-level essays written by Pakistani ESL students are analyzed using both Spearman's correlation and ridge regression to address the issues of non-normality and multicollinearity in the data. The narrativity indices analysed are PCVERBz (verb cohesion), PCNARz (narrative cohesion), and SMCAUSv/SMCAUSsa (causal reasoning).

This study aims to address a significant gap in the literature on how narrativity features impact writing quality in the Pakistani ESL context. In this area, corpus-based writing research is under-represented. While, in contrast, lexical and grammatical features have been extensively studied, the role of narrativity in writing performance is still, to some degree, not fully explored, particularly within the context of the assessment of high stakes, in addition to forming a better understanding of these relationships to create stronger targeted instructional interventions and the capability to engineer more refined automated assessment tools.

According to this study, excessive verb cohesion (repetition) exhibits a negative correlation with writing scores, whereas causal reasoning and narrative cohesion present a positive correlation with coherence and task performance. These insights have theoretical contributions to prove and extend the research on discourse-level features of writing, as well as practical implications for ESL pedagogy and automated scoring systems. Ultimately, the study aims to inform activities designed to enhance English for Academic Purposes (EAP) assessment and instruction in multilingual academic settings.

## **2. Literature Review**

Narrativity studies have focused on the features of narrative and their potential applications in the field of ESL academic writing. Many researchers believe that discourse-related features, such as narrative cohesion, verb cohesion, and causal reasoning, play a crucial role in assessing writing skills. Coherent writing—essential for academic success—can be achieved through specific types of structures and functions of text, such as the theme-rheme system, which organizes information in a logically ordered manner (Haider, 2022). Corpus-based studies in the Pakistani ESL context have indicated, among other things, that lexical inappropriacy, coherence problems, and limited exposure to academic genres contribute to the erosion of the development of good writing skills (Gill, Kausar, & Haider, 2022; Ramzan,

Mushtaq, & Ashraf, n.d.). Recent meta-analytic studies confirm that corpus-based instruction has greatly benefited ESL writing worldwide (Ngo & Chen, 2024), thereby emphasizing the pedagogical significance of linguistic feature analysis in addressing learner weaknesses.

### **1. Writing Quality in ESL Contexts**

Writing quality in ESL contexts is usually evaluated on criteria such as coherence and cohesion, lexical resource, grammar, and task response, which serve as the core of assessment instruments such as IELTS and TOEFL. Cohesion serves as the vehicle for achieving logical flow from one idea to another. It is an oft-highlighted area of difficulty for ESL learners in both Pakistani and non-Pakistani contexts (Crossley et al., 2014). For example, Liu and Braine (2005) found that Chinese learners relied heavily on local cohesive devices, which constrained the development of global coherence. In contrast, Behnam and Golpour (2014) found that Iranian students struggled with transitions at the paragraph level, primarily due to a lack of discourse training. In the Pakistani context, students typically produce descriptions rather than arguments due to the overuse of linear thematic structures (Haider, 2022). However, corpus-based methodologies have been found to help learners employ more advanced lexico-grammatical features that enhance cohesion and coherence (Ngo & Chen, 2024). In conjunction with appraisal theory, the importance of cohesion for persuasive academic texts has also been demonstrated (Gill, Kausar, & Haider, 2022).

### **3. Past Research on Narrativity and Writing Performance**

Despite a small quantity of research on certain individual narrativity elements, much less research has thoughtfully explored the interaction of these along writing performance lines, and least of all in the Pakistani context. Haider (2022) found that Pakistani undergraduates tend to overuse pronominal modifiers and constant thematic progression, which has a demeaning effect that detracts from argument development. According to Qasim et al. (n.d.), there is also excessive use of additive connectors, which causes a lack of cohesion. Similarly, Cheung (2023) emphasized that the overuse of conjunctions, such as "and" or "then," by young writers diminished textual complexity and resulted in lower writing scores. In the broader computational space of Pakistani ESL corpora, this reliance on simpler, repetitive items matches patterns where foundational lexical frequency profiles limit text-level complexity (Bano et al., 2025). Furthermore, basic accuracy errors have been shown to constrain advanced sentence generation, driving students to adopt shorter text segments to mitigate systemic error rates (Mahmood et al., 2026). More importantly, narrativity features appear to be associated with comprehension. Villarroel-Henríquez and Bruna-Jofre (2024) found that research-rich teacher discourse enhanced student understanding, suggesting that a similar mechanism may be at work in the writing arena.

Studies have been conducted to examine general text features, such as cohesion and readability, using Coh-Metrix (Crossley et al., 2011). However, few have focused on the connection between narrativity indices and performance scores in standardized writing tasks. This study, however, stands apart because it implements a set of different narrativity features (PCVERBz, PCNARz, SMCAUSvlsa) and maps them directly onto IELTS-aligned band scores from both correlation and ridge regression modelling, making an unprecedented contribution to the Pakistani ESL research space.

### **4. Gaps in Existing Research**

While the interest in automated writing analysis and corpus-informed pedagogy is on the rise, some gaps exist. They include:

1. **Weak focus on narrativity in Pakistani ESL contexts:** Most studies at the local level concentrate strictly on syntax and vocabulary. While recent local literature has successfully utilized Coh-Metrix to analyze the constraints of error patterns on syntactic complexity (Mahmood et al., 2026) and the predictive value of word-concreteness indexes (Bano et al., 2025) , multi-feature interactions mapping narrativity directly onto macro-level writing criteria remain heavily underrepresented.
2. **Lack of cross-cultural comparison:** Findings from other ESL populations (e.g., Iranian, Chinese, and Arab learners) are hardly used to draw parallels with Pakistani data.
3. **Insufficient integration of multiple narrativity features:** Narrativity research often analyses cohesion elements in isolation rather than considering their combined effect on writing quality.
4. **Alignment with high-stakes assessments is often neglected:** very few investigations seek to operationally relate narrativity indices to components of writing that are core to grading systems, such as the IELTS or TOEFL.

## 5. Need for Integrated Analysis

There is a pressing need for multi-feature analysis that examines how narrativity traits interact to influence writing performance, particularly in settings where rubric-orientated assessment programs are implemented. Previously, authors have employed separatist approaches with cohesion and coherence styles or merely scratched the surface concerning wording mechanisms. Conversely, this dissertation examines how verb cohesion, narrative cohesion, and causal reasoning variables compare in terms of Coh-Metrix scores and writing scores, which align with the IELTS. This marks the transition point, where theoretical linguistics come together, practically paving the way for EAP instruction, teacher training, and, at a later stage, the development of automated scoring.

## Research Objectives:

1. To investigate the impact of narrativity features on Pakistani ESL learners' writing and their performance (verb cohesion, narrative cohesion, and causal reasoning).
2. To analyze the relationship between narrativity features and the four IELTS-style writing band scores: Coherence and cohesion (CC), Lexical Resource (LR), Grammar (GR), and Task Response (TR).
3. To analyze the difficulties the Pakistani ESL students go through when they apply narrativity features, especially when performing academic writing tasks.
4. To suggest corpus-driven pedagogical activities with the intent of enhancing the application of narrativity features in favor of ESL students' writing performance.

## Research Questions:

1. How do narrativity features (verb cohesion, narrative cohesion, and causal reasoning) impact the writing performance of Pakistani ESL students in IELTS-style assessments?
2. What is the relationship between verb cohesion, narrative cohesion, and causal reasoning and the IELTS-style writing band scores (Coherence and Cohesion, Lexical Resource, Grammar, and Task Response) of Pakistani ESL students?
3. What challenges do Pakistani ESL students face in incorporating narrativity features, and how can these challenges be addressed to improve their writing?
4. What corpus-based pedagogical activities can be designed to help Pakistani ESL students improve their use of narrativity features and enhance their writing quality?

### 3. Methodology

#### Research Design

A quantitative, correlational approach has been adopted here in the context of corpus-based linguistic analyses. The study's primary objective has been to assess the role of narrativity features, specifically verb and narrative cohesion, as well as causal reasoning, in determining writing performance among Pakistani ESL undergraduate students.

#### Participants and Data Collection

A corpus of 100 essays on argumentative topics written by ESL undergraduate students from different universities in Punjab, Pakistan, has been analyzed. This corpus profiling and regional selection pattern align with parallel quantitative paradigms in the region, which have validated the utility of utilizing 100-essay public and private university sample frameworks across Punjab for computational analysis (Bano et al., 2025; Mahmood et al., 2026). The participants have been drawn from different linguistic, disciplinary, and gender backgrounds.

#### Essay Scoring

For performance evaluation, the system scores essays using locally developed automated scoring software that claims to replicate the IELTS Task 2 band descriptors. Four different components of the system have been set against the parameters that include:

1. Coherence and Cohesion (CC)
2. Lexical Resource (LR)
3. Grammatical Range and Accuracy (GR)
4. Task Response (TR)

#### Linguistic Feature Extraction

Linguistic features are extracted using Coh-Metrix, a computational tool for text analysis that gives more than 100 indices related to cohesion, narrativity, syntax, and readability (Graesser et al., 2004). For this study, a total of 26 narrativity-related features have been extracted and categorized under discourse structure and cohesion. These measures include referential cohesion, temporal cohesion, narrativity, and causal connectives.

From the preliminary correlation analyses run on all 26 features, the following indices have arisen as statistically significant predictors of performance (either positively or negatively):

PCVERBz: Verb cohesion (z-scored)

PCNARz: Narrative cohesion (z-scored)

SMCAUSv / SMCAUSlsa: Causal reasoning via verb overlap and latent semantic analysis

Following the evidence of the predictive relationships these features have with IELTS-style band scores across several components, they will be submitted for further analysis.

#### Data Analysis Techniques

The following statistical strategies have been seen executed:

1. Spearman's rank correlation is used because the scores of writing and feature indices are not normally distributed.
2. Ridge regression analysis is used to model the relationship between narrativity features and writing band scores, thereby addressing issues of multicollinearity and overfitting.
3. The  $R^2$  score, cross-validation scores, and standardized coefficients have been used to measure model performances. This mixed statistical layout allows for the modelling of both predictive and correlational relationships.

## Ethical Considerations

The data of participants have been anonymized for privacy before commencing analysis. All formalities for data collection have been approved at an institutional level, and no information that can lead to the identification of any participant has been recorded. The essays themselves have been gathered for academic purposes with the well-informed consent of the individuals.

## 4. Data Analysis & Results

### Descriptive Statistics

| <i>Feature</i>                   | <i>Mean</i> | <i>Std Dev</i> | <i>Min</i> | <i>Max</i> | <i>Median</i> | <i>1st<br/>Quartile</i> | <i>3rd<br/>Quartile</i> |
|----------------------------------|-------------|----------------|------------|------------|---------------|-------------------------|-------------------------|
| <i>PCNARz</i>                    | -0.17       | 0.89           | -2.04      | 2.27       | -0.15         | -0.70                   | 0.48                    |
| <i>PCNARp</i>                    | 43.64       | 28.12          | 2.28       | 98.84      | 44.83         | 19.22                   | 68.08                   |
| <i>PCREFz</i>                    | 0.50        | 1.44           | -2.20      | 4.84       | 0.37          | -0.48                   | 1.29                    |
| <i>PCREFp</i>                    | 64.43       | 28.58          | 1.39       | 100.00     | 70.19         | 42.86                   | 89.25                   |
| <i>PCVERBz</i>                   | 0.24        | 1.57           | -3.59      | 6.70       | 0.31          | -0.76                   | 1.05                    |
| <i>PCVERBp</i>                   | 56.17       | 30.46          | 0.00       | 100.00     | 61.41         | 28.17                   | 84.61                   |
| <i>PCDCz</i>                     | 1.13        | 1.70           | -2.41      | 7.81       | 1.20          | 0.40                    | 1.85                    |
| <i>PCDCp</i>                     | 73.55       | 24.88          | 1.92       | 100.00     | 79.94         | 65.17                   | 92.22                   |
| <i>CRFAOI</i>                    | 0.58        | 0.23           | 0.05       | 1.00       | 0.58          | 0.39                    | 0.75                    |
| <i>CRFAOa</i>                    | 0.53        | 0.21           | 0.14       | 1.00       | 0.52          | 0.36                    | 0.68                    |
| <i>SMCAUSv</i>                   | 30.85       | 15.53          | 0.00       | 76.36      | 28.85         | 20.41                   | 41.19                   |
| <i>SMCAUSvp</i>                  | 47.32       | 17.07          | 3.13       | 87.27      | 47.73         | 34.33                   | 58.09                   |
| <i>SMCAUSr</i>                   | 0.67        | 0.81           | 0.00       | 11.00      | 0.55          | 0.25                    | 0.80                    |
| <i>SMCAUSlsa</i>                 | 0.12        | 0.09           | 0.02       | 0.56       | 0.10          | 0.07                    | 0.15                    |
| <i>SMCAUSwn</i>                  | 0.53        | 0.16           | 0.18       | 0.84       | 0.52          | 0.43                    | 0.64                    |
| <i>SMINTEp</i>                   | 19.62       | 13.80          | 0.00       | 57.88      | 16.29         | 9.05                    | 28.67                   |
| <i>SMINTER</i>                   | 1.45        | 1.35           | 0.09       | 9.00       | 1.17          | 0.75                    | 1.83                    |
| <i>PCTEMPz</i>                   | -0.16       | 1.15           | -3.16      | 4.07       | -0.23         | -0.71                   | 0.41                    |
| <i>PCTEMPp</i>                   | 46.23       | 27.21          | 0.08       | 100.00     | 42.86         | 24.20                   | 66.28                   |
| <i>WRDVERB</i>                   | 118.25      | 22.18          | 41.15      | 183.52     | 116.92        | 104.95                  | 128.38                  |
| <i>DRGERUND</i>                  | 18.84       | 14.12          | 0.00       | 64.82      | 17.07         | 8.59                    | 26.06                   |
| <i>DRINF</i>                     | 14.60       | 9.84           | 0.00       | 42.76      | 13.33         | 7.82                    | 20.49                   |
| <i>CNCAll</i>                    | 100.34      | 24.31          | 37.93      | 156.86     | 98.32         | 84.30                   | 117.26                  |
| <i>CNCCaus</i>                   | 33.09       | 15.53          | 2.68       | 73.17      | 30.77         | 21.39                   | 43.94                   |
| <i>CNCLogic</i>                  | 50.47       | 17.85          | 10.72      | 90.09      | 48.19         | 37.93                   | 62.50                   |
| <i>CNCTemp</i>                   | 12.86       | 9.60           | 0.00       | 32.41      | 11.66         | 6.15                    | 18.24                   |
| <i>CC</i>                        | 4.47        | 0.62           | 2.79       | 6.75       | 4.60          | 3.95                    | 4.60                    |
| <i>GR</i>                        | 4.77        | 0.90           | 2.39       | 7.49       | 4.95          | 4.09                    | 5.26                    |
| <i>LR</i>                        | 5.09        | 0.69           | 3.84       | 7.40       | 5.24          | 4.67                    | 5.49                    |
| <i>TR</i>                        | 5.10        | 0.73           | 2.98       | 7.45       | 5.11          | 4.48                    | 5.58                    |
| <i>Predicted<br/>Total Score</i> | 5.09        | 0.85           | 3.12       | 7.54       | 5.21          | 4.36                    | 5.56                    |

Descriptive statistics for narrativity-related features show that PCDCz (SD = 1.70) and PCVERBz (SD = 1.57) exhibit the highest variability across student essays, indicating that there are differences in the presence and use of these elements. On the contrary, SMCAUSlsa (SD = 0.09) and SMCAUSwn (SD = 0.16) show more consistent patterns, indicating causal markers are consistently present. These features, together with DRGERUND and CNCTemp, also show minimum values of zero, implying that some essays completely lack these narrative elements.

Looking at the band scores of IELTS, CC stands at the lowest position with a mean score of 4.47 and the least variation (SD = 0.62), which indicates it is a common area of weakness among students, with a restricted range of dispersion. Grammar has the maximum variation (SD = 0.90), which means extremes in grammatical proficiencies. Lexical resource and task response rates are higher, with both averaging nearly 5.1, indicating closer alignment with the acquired areas from the level's side in performance. Overall, the majority of the scores fall within the intermediate band, roughly between 4.5 and 5.5, indicating that the vast majority of performances are concentrated in this zone.

### Normality Test Results

| <i>Variable</i>              | <i>Test Statistic</i> | <i>p-value</i> | <i>Normal Distribution?</i> |
|------------------------------|-----------------------|----------------|-----------------------------|
| <b>PCNARz</b>                | 0.970                 | 0.021          | No*                         |
| <b>PCNARp</b>                | 0.974                 | 0.044          | No*                         |
| <b>PCREFz</b>                | 0.974                 | 0.049          | No*                         |
| <b>PCREFp</b>                | 0.897                 | 9.92e-07       | No                          |
| <b>PCVERBz</b>               | 0.946                 | 4.76e-04       | No                          |
| <b>PCVERBp</b>               | 0.941                 | 2.21e-04       | No                          |
| <b>PCDCz</b>                 | 0.935                 | 1.03e-04       | No                          |
| <b>PCDCp</b>                 | 0.868                 | 5.77e-08       | No                          |
| <b>CRFAO1</b>                | 0.978                 | 0.092          | Yes                         |
| <b>CRFAOa</b>                | 0.988                 | 0.475          | Yes                         |
| <b>SMCAUSv</b>               | 0.981                 | 0.153          | Yes                         |
| <b>SMCAUSvp</b>              | 0.991                 | 0.777          | Yes                         |
| <b>SMCAUSr</b>               | 0.392                 | 1.88e-18       | No                          |
| <b>SMCAUSlsa</b>             | 0.796                 | 1.87e-10       | No                          |
| <b>SMCAUSwn</b>              | 0.981                 | 0.156          | Yes                         |
| <b>SMINTEp</b>               | 0.934                 | 8.83e-05       | No                          |
| <b>SMINTER</b>               | 0.530                 | 2.45e-16       | No                          |
| <b>PCTEMPz</b>               | 0.360                 | 6.87e-19       | No                          |
| <b>PCTEMPp</b>               | 0.956                 | 0.002          | No                          |
| <b>WRDVERB</b>               | 0.974                 | 0.042          | No*                         |
| <b>DRGERUND</b>              | 0.942                 | 2.40e-04       | No                          |
| <b>DRINF</b>                 | 0.959                 | 0.004          | No                          |
| <b>CNCAI1</b>                | 0.990                 | 0.670          | Yes                         |
| <b>CNCCaus</b>               | 0.970                 | 0.022          | No*                         |
| <b>CNCLogic</b>              | 0.995                 | 0.983          | Yes                         |
| <b>CNCTemp</b>               | 0.937                 | 1.20e-04       | No                          |
| <b>CC</b>                    | 0.818                 | 9.38e-10       | No                          |
| <b>GR</b>                    | 0.903                 | 2.01e-06       | No                          |
| <b>LR</b>                    | 0.845                 | 8.15e-09       | No                          |
| <b>TR</b>                    | 0.969                 | 0.018          | No*                         |
| <b>Predicted Total Score</b> | 0.935                 | 1.03e-04       | No                          |

The variables CRFAO1, CRFAOa, SMCAUSv, SMCAUSvp, SMCAUSwn, CNCAI1, and CNCLogic exhibit normality ( $p > 0.05$ ) and thus qualify for parametric analyses. On the other hand, many variables severely violate normality ( $p < 0.001$ ): PCVERBz, PCDCz, SMCAUSlsa, DRGERUND, IELTS primary scoring criteria of CC, GR, LR, Predicted Total Score, etc. Such a severe violation implies that these narrative features and performance scores are significantly skewed or exhibit non-normal patterns for these criteria. Other variables, such as PCNARz, PCREFz, CNCCaus, and TR, show a moderate violation ( $p < 0.05$ ) and may pose a concern for statistical modelling. Thus, most key performance and linguistic variables are far from achieving normality assumptions.

Given that 77% of variables violate the normality assumption, Spearman's rank correlation will be used instead of Pearson's correlation to investigate the relationships between narrativity features and band scores.

### Spearman's Rank Correlation Results

| FEATURE   | CC                  | GR                  | LR                  | TR                  | PREDICTED<br>TOTAL<br>SCORE |
|-----------|---------------------|---------------------|---------------------|---------------------|-----------------------------|
| PCNARZ    | -0.330<br>(p=0.001) | -0.314<br>(p=0.001) | -0.302<br>(p=0.002) | -0.372<br>(p=0.000) | -0.345<br>(p=0.000)         |
| PCNARP    | -0.331<br>(p=0.001) | -0.316<br>(p=0.001) | -0.304<br>(p=0.002) | -0.373<br>(p=0.000) | -0.346<br>(p=0.000)         |
| PCREFZ    | -0.328<br>(p=0.001) | -0.359<br>(p=0.000) | -0.338<br>(p=0.001) | -0.357<br>(p=0.000) | -0.350<br>(p=0.000)         |
| PCREFP    | -0.326<br>(p=0.001) | -0.357<br>(p=0.000) | -0.336<br>(p=0.001) | -0.355<br>(p=0.000) | -0.349<br>(p=0.000)         |
| PCVERBZ   | -0.364<br>(p=0.000) | -0.434<br>(p=0.000) | -0.432<br>(p=0.000) | -0.355<br>(p=0.000) | -0.384<br>(p=0.000)         |
| PCVERBP   | -0.364<br>(p=0.000) | -0.434<br>(p=0.000) | -0.431<br>(p=0.000) | -0.356<br>(p=0.000) | -0.384<br>(p=0.000)         |
| PCDCZ     | -0.034<br>(p=0.736) | -0.056<br>(p=0.579) | -0.067<br>(p=0.506) | 0.020<br>(p=0.845)  | -0.013<br>(p=0.896)         |
| PCDCP     | -0.033<br>(p=0.741) | -0.056<br>(p=0.583) | -0.067<br>(p=0.506) | 0.019<br>(p=0.849)  | -0.013<br>(p=0.895)         |
| CRFAO1    | -0.194<br>(p=0.053) | -0.199<br>(p=0.047) | -0.186<br>(p=0.064) | -0.244<br>(p=0.014) | -0.218<br>(p=0.030)         |
| CRFAOA    | -0.122<br>(p=0.227) | -0.145<br>(p=0.150) | -0.125<br>(p=0.214) | -0.147<br>(p=0.145) | -0.143<br>(p=0.155)         |
| SMCAUSV   | 0.274<br>(p=0.006)  | 0.301<br>(p=0.002)  | 0.267<br>(p=0.007)  | 0.182<br>(p=0.070)  | 0.244<br>(p=0.015)          |
| SMCAUSVP  | 0.190<br>(p=0.058)  | 0.193<br>(p=0.055)  | 0.182<br>(p=0.070)  | 0.142<br>(p=0.159)  | 0.165<br>(p=0.101)          |
| SMCAUSR   | -0.212<br>(p=0.034) | -0.246<br>(p=0.014) | -0.227<br>(p=0.023) | -0.130<br>(p=0.199) | -0.204<br>(p=0.042)         |
| SMCAUSLSA | -0.231<br>(p=0.021) | -0.316<br>(p=0.001) | -0.281<br>(p=0.005) | -0.238<br>(p=0.017) | -0.259<br>(p=0.009)         |
| SMCAUSWN  | -0.024<br>(p=0.813) | -0.087<br>(p=0.387) | -0.089<br>(p=0.379) | -0.031<br>(p=0.757) | -0.037<br>(p=0.713)         |
| SMINTEP   | 0.167<br>(p=0.097)  | 0.132<br>(p=0.191)  | 0.167<br>(p=0.097)  | 0.007<br>(p=0.941)  | 0.076<br>(p=0.453)          |
| SMINTER   | -0.103<br>(p=0.306) | -0.052<br>(p=0.610) | -0.096<br>(p=0.343) | 0.050<br>(p=0.619)  | -0.004<br>(p=0.972)         |
| PCTEMPZ   | -0.153<br>(p=0.129) | -0.193<br>(p=0.055) | -0.175<br>(p=0.082) | -0.080<br>(p=0.431) | -0.117<br>(p=0.246)         |
| PCTEMPP   | -0.155<br>(p=0.123) | -0.195<br>(p=0.051) | -0.177<br>(p=0.078) | -0.081<br>(p=0.421) | -0.119<br>(p=0.237)         |
| WRDVERB   | 0.215<br>(p=0.032)  | 0.297<br>(p=0.003)  | 0.272<br>(p=0.006)  | 0.195<br>(p=0.051)  | 0.225<br>(p=0.024)          |
| DRGERUND  | 0.264<br>(p=0.008)  | 0.339<br>(p=0.001)  | 0.320<br>(p=0.001)  | 0.305<br>(p=0.002)  | 0.309<br>(p=0.002)          |
| DRINF     | 0.109<br>(p=0.282)  | 0.150<br>(p=0.137)  | 0.148<br>(p=0.141)  | 0.083<br>(p=0.410)  | 0.116<br>(p=0.251)          |
| CNCALL    | -0.020              | 0.038               | -0.016              | 0.085               | 0.059                       |

|          |           |           |           |           |           |
|----------|-----------|-----------|-----------|-----------|-----------|
|          | (p=0.841) | (p=0.710) | (p=0.878) | (p=0.402) | (p=0.561) |
| CNCCAUS  | -0.045    | -0.073    | -0.059    | 0.001     | -0.031    |
|          | (p=0.654) | (p=0.470) | (p=0.561) | (p=0.990) | (p=0.758) |
| CNCLOGIC | 0.060     | -0.009    | -0.013    | 0.039     | 0.025     |
|          | (p=0.553) | (p=0.925) | (p=0.897) | (p=0.697) | (p=0.805) |
| CNCTEMP  | 0.050     | 0.126     | 0.059     | 0.128     | 0.105     |
|          | (p=0.622) | (p=0.212) | (p=0.558) | (p=0.205) | (p=0.298) |

**Note:**  $\rho$  = Spearman's rho correlation coefficient. Correlations above  $\pm 0.3$  are considered moderate and practically meaningful in educational contexts.

The statistical evaluation reveals several extremely significant relationships ( $p < 0.001$ ) between the features of narrativity and IELTS band scores. It is important to note that these features related to verbs, PCVERBz, PCVERBp, PCREFz, and PCREFp, are all negatively correlated with scores in each band ( $r = -0.302$  and  $-0.434$ ), which in turn implies high cohesion brought about by repetitive use of verbs to be a signifier of poor writing, as a result of less lexical extenderness and fleetingness. Similarly, a negative relationship between PCNARz/PCNARp and band scores also supports this trend.

Conversely, aspects related to synthetic complexity, such as DRGERUND (gerund density), positively correlate highly with Grammar, Lewis Resource, Task, and Response, as well as the Predicted Total Score ( $r = 0.305$  to  $0.339$ ). SMCAUSv is also positively correlated with Grammar ( $r = 0.301$ ), indicating that the use of causal connectors is associated with better performance.

Nonetheless, other features of causal cohesion, such as SMCAUSlsa and SMCAUSr, show negative correlations (e.g., SMCAUSlsa with GR,  $r = -0.316$ ), highlighting the detailed importance of causal reasoning in the assessment of scholarly writing. Other positive correlations are also found at the  $p < 0.05$  level between WRDVERB, SMCAUSv, and DRGERUND, as well as different elements of the scoring. On the other hand, CRFAO1, SMCAUSr and SMCAUSlsa have negative associations.

The other variables, such as PCDCz/PCDCp, CRFAOa, SMCAUSwn, and most connective measures, indicate a weak or non-significant correlation, implying that they had little influence on writing performance. By and large, closest relations to the quality of writing are observed with verb-related narrativity features. Although verbs are high by their cohesion scores, as well as high by either verb diversity or the use of gerunds in place of "simple" verbs, they suggest lower scores at the advanced level of syntactic and lexical control. Causal cohesion is effective but with a more complicated and transactional relationship in connection with performance outcomes.

### Group Comparison

The hundred essays were divided into two groups based on the Total Score Predicted: a scoring group with scores  $\geq 5$  (54 essays) and a low-scoring group with scores  $< 5$  (46 essays). Statistical testing has been selected based on the normality of the features. Features that have violated the normality assumption are subjected to a Mann-Whitney U test, while those that meet the normality criteria are subjected to an independent t-test.

| <i>Feature</i>   | <i>Test Used</i> | <i>Test Statistic</i> | <i>p-value</i> | <i>Significant?</i> |
|------------------|------------------|-----------------------|----------------|---------------------|
| <b>PCNARz</b>    | Mann-Whitney U   | 873.00                | 0.011          | <b>Yes</b>          |
| <b>PCNARp</b>    | Mann-Whitney U   | 871.50                | 0.010          | <b>Yes</b>          |
| <b>PCREFz</b>    | Mann-Whitney U   | 854.00                | 0.007          | <b>Yes</b>          |
| <b>PCREFp</b>    | Mann-Whitney U   | 855.50                | 0.008          | <b>Yes</b>          |
| <b>PCVERBz</b>   | Mann-Whitney U   | 907.00                | 0.021          | <b>Yes</b>          |
| <b>PCVERBp</b>   | Mann-Whitney U   | 908.00                | 0.021          | <b>Yes</b>          |
| <b>PCDCz</b>     | Mann-Whitney U   | 1284.00               | 0.774          | No                  |
| <b>PCDCp</b>     | Mann-Whitney U   | 1282.00               | 0.785          | No                  |
| <b>CRFAO1</b>    | t-test           | -1.76                 | 0.082          | No                  |
| <b>CRFAOa</b>    | t-test           | -1.48                 | 0.142          | No                  |
| <b>SMCAUSv</b>   | t-test           | 2.47                  | 0.015          | <b>Yes</b>          |
| <b>SMCAUSvp</b>  | t-test           | 2.32                  | 0.022          | <b>Yes</b>          |
| <b>SMCAUSr</b>   | Mann-Whitney U   | 1076.00               | 0.252          | No                  |
| <b>SMCAUSlsa</b> | Mann-Whitney U   | 1087.00               | 0.285          | No                  |
| <b>SMCAUSwn</b>  | t-test           | 0.42                  | 0.675          | No                  |
| <b>SMINTEp</b>   | Mann-Whitney U   | 1424.50               | 0.208          | No                  |
| <b>SMINTER</b>   | Mann-Whitney U   | 1215.00               | 0.854          | No                  |
| <b>PCTEMPz</b>   | Mann-Whitney U   | 1208.50               | 0.819          | No                  |
| <b>PCTEMPp</b>   | Mann-Whitney U   | 1205.50               | 0.803          | No                  |
| <b>WRDVERB</b>   | Mann-Whitney U   | 1469.00               | 0.117          | No                  |
| <b>DRGERUND</b>  | Mann-Whitney U   | 1466.00               | 0.122          | No                  |
| <b>DRINF</b>     | Mann-Whitney U   | 1302.50               | 0.678          | No                  |
| <b>CNCAI</b>     | t-test           | 0.21                  | 0.832          | No                  |
| <b>CNCCaus</b>   | t-test           | 0.26                  | 0.796          | No                  |
| <b>CNCLogic</b>  | t-test           | 0.42                  | 0.673          | No                  |
| <b>CNCTemp</b>   | Mann-Whitney U   | 1291.00               | 0.737          | No                  |

Several cohesion features demonstrate statistically significant differences ( $p < 0.05$ ) between high- and low-scoring groups. Namely, group differences are observable for some selected statistical measures of narrative cohesion: PCNARz ( $p = 0.011$ ) and PCNARp ( $p = 0.010$ ), as well as reference cohesion: PCREFz ( $p = 0.007$ ) and PCREFp ( $p = 0.008$ ). Differences between groups are also statistically significant for verb cohesion indicators PCVERBz and PCVERBp (both  $p = 0.021$ ). Considering causal cohesion, SMCAUSv ( $p = 0.015$ ) and SMCAUSvp ( $p = 0.022$ ) diverge significantly. On the other hand, features that produce no statistically significant group differences include specific deep cohesion measures (PCDCz, PCDCp); first-person pronoun ratios (CRFAO1, CRFAOa); most causal cohesion features (SMCAUSr, SMCAUSlsa, SMCAUSwn); intentional cohesion (SMINTEp, SMINTER); temporal cohesion (PCTEMPz, PCTEMPp); verb diversity and density (WRDVERB, DRGERUND, DRINF); and connective measures (CNCAI, CNCCaus, CNCLogic, CNCTemp).

The results indicate that 31% of narrativity features differ statistically between high- and low-scoring groups. These features are primarily centered around the basic cohesion features—narrative, reference, and verb cohesion, while also incorporating some causal cohesion features. This could indicate that rather than higher-level cohesive devices, it is the basic cohesive features that separate better from low-performing student essays. Brazenly, the very difference in verb cohesion lends support to the previous correlational result by stating that more verb repetition might be a sign of lower-level writing.

## Multiple Regression Analysis

| <i>Writing Descriptor</i>            | <i>R-squared</i> | <i>Adjusted squared</i> | <i>R-</i> | <i>F-statistic</i> | <i>p-value</i> | <i>Model Significance</i> |
|--------------------------------------|------------------|-------------------------|-----------|--------------------|----------------|---------------------------|
| <i>Coherence &amp; Cohesion (CC)</i> | 0.611            | 0.473                   |           | 4.411              | 2.92e-07       | Highly Significant        |
| <i>Lexical Resource (LR)</i>         | 0.626            | 0.492                   |           | 4.691              | 9.50e-08       | Highly Significant        |
| <i>Grammar (GR)</i>                  | 0.585            | 0.437                   |           | 3.952              | 1.94e-06       | Highly Significant        |
| <i>Task Response (TR)</i>            | 0.498            | 0.319                   |           | 2.787              | 0.000319       | Significant               |

### Coherence and Cohesion (CC) - Best Model Performance

The regression analysis has identified several important predictors of CC scores ( $p < 0.05$ ). PCVERBz is a strong negative predictor, which means that the larger the value given for the z-score measure of verb cohesion, the lower the CC score ( $\beta = -0.441$ ,  $p = 0.010$ ). Conversely, PCVERBp has a positive relationship, so the higher the percentile rank of verb cohesion, the higher the CC performance will be to some extent ( $\beta = 0.010$ ,  $p = 0.048$ ). SMCAUSlsa is a strong positive predictor, which gives a greater emphasis on causal cohesion through LSA in enhancing CC scoring power ( $\beta = 3.856$ ,  $p = 0.013$ ). Notably, CNCAI1 occurs as a negative predictor, implying that an increase in the number of connectives may correspond to less coherence or under coherence ( $\beta = -0.007$ ,  $p = 0.050$ ). In summary, the model explains 61.1% of the variance in CC scores.

### Lexical Resource (LR) - The Most Powerful Predictive Model

The regression analysis has found significant predictors of Lexical Resource (LR) scores ( $p < 0.05$ ). PCVERBz is also the strongest negative one ( $\beta = -0.525$ ,  $p = 0.005$ ), indicating that too much verb cohesion diminishes lexical sophistication. On the other hand, SMCAUSlsa is the strongest positive predictor ( $\beta = 5.463$ ,  $p = 0.001$ ), indicating that causal cohesion, aided by LSA, significantly increases lexical resource scores. Near significant predictors include PCNARz ( $\beta = -0.791$ ,  $p = 0.053$ ), with a marginally negative effect of narrative cohesion on LR, and WRDVERB ( $\beta = 0.006$ ,  $p = 0.092$ ), showing a marginally positive effect of verb diversity. CNCAI1 ( $\beta = -0.007$ ,  $p = 0.060$ ) indicates a marginally negative effect of the use of all connectives. Overall, the model accounts for 62.6% of the variance in LR scores.

### Grammar (GR) - Strong Predictive Performance

The regression reveals significant predictors of GR scores ( $p < 0.05$ ). PCVERBz is a strong negative predictor ( $\beta = -0.602$ ,  $p = 0.018$ ), suggesting that repetitions of verb patterns limit grammar. SMCAUSlsa is a strong positive predictor ( $\beta = 6.166$ ,  $p = 0.009$ ), indicating a refined causal reasoning ability that enhances grammar range. CNCAI1 ( $\beta = -0.010$ ,  $p = 0.071$ ) is an almost significant predictor, indicating a slight negative influence of the use of all connectives. The model explains 58.5% of the variance in GR scores.

### Task Response (TR) - Moderate Predictive Power

The regression analysis identifies significant predictors of TR scores ( $p < 0.05$ ). PCVERBz is a strong negative predictor ( $\beta = -0.562$ ,  $p = 0.013$ ), indicating that verb cohesion patterns are detrimental to task fulfilment. SMCAUSlsa is the strongest positive predictor ( $\beta = 6.586$ ,  $p =$

0.002), underscoring the importance of causal reasoning in enhancing task response. The model explains 49.8% of the variance in TR scores.

### Across-Model Patterns

The analysis helps identify predictors that are consistent across all models. PCVERBz (Verb Cohesion Z-score) is the one significant negative predictor in each model, showing that excessive repetition of verbs consistently predicts lower writing quality. CNCAII (All Connectives) also shows a negative effect throughout the models, indicating that the overuse of connectives may be a hallmark of less sophisticated writing. Meanwhile, SMCAUSlsa (Causal Cohesion LSA) is consistently positive and is the strongest predictor in all models, indicating that sophisticated causal reasoning is critical for all aspects of writing quality.

### Model Diagnostics:

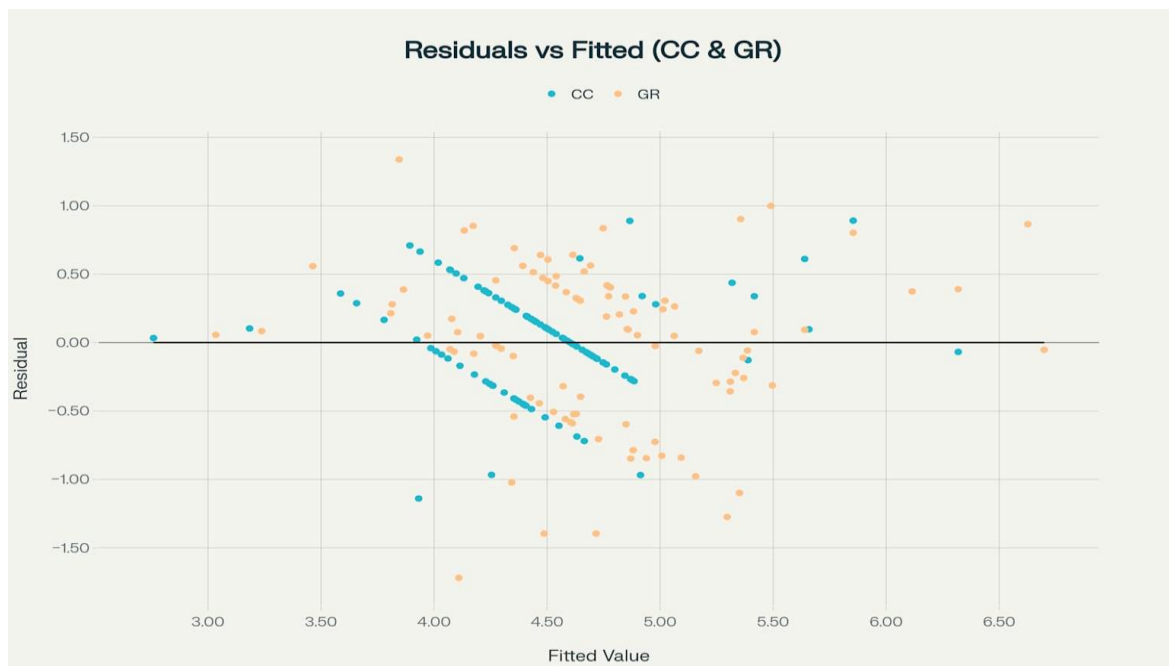
#### Variance Inflation Factor (VIF) Analysis:

| <i>Feature</i>   | <i>VIF Value</i> | <i>Multicollinearity Level</i> |
|------------------|------------------|--------------------------------|
| <b>PCDCz</b>     | 188.51           | <b>Extreme</b>                 |
| <b>PCDCp</b>     | 187.42           | <b>Extreme</b>                 |
| <b>PCNARz</b>    | 44.54            | <b>Severe</b>                  |
| <b>PCNARp</b>    | 39.87            | <b>Severe</b>                  |
| <b>SMCAUSr</b>   | 34.60            | <b>Severe</b>                  |
| <b>CNCCaus</b>   | 31.99            | <b>Severe</b>                  |
| <b>PCVERBz</b>   | 30.94            | <b>Severe</b>                  |
| <b>PCVERBp</b>   | 30.87            | <b>Severe</b>                  |
| <b>CNCLogic</b>  | 23.72            | <b>High</b>                    |
| <b>PCREFz</b>    | 17.04            | <b>High</b>                    |
| <b>SMINTER</b>   | 16.34            | <b>High</b>                    |
| <b>PCREFp</b>    | 15.89            | <b>High</b>                    |
| <b>PCTEMPz</b>   | 12.45            | <b>High</b>                    |
| <b>PCTEMPp</b>   | 12.41            | <b>High</b>                    |
| <b>SMCAUSv</b>   | 8.92             | <b>Moderate</b>                |
| <b>SMCAUSvp</b>  | 8.87             | <b>Moderate</b>                |
| <b>CNCAII</b>    | 7.23             | <b>Moderate</b>                |
| <b>WRDVERB</b>   | 6.78             | <b>Moderate</b>                |
| <b>SMINTEp</b>   | 5.94             | <b>Moderate</b>                |
| <b>DRGERUND</b>  | 4.89             | <b>Moderate</b>                |
| <b>CNCTemp</b>   | 4.12             | <b>Moderate</b>                |
| <b>DRINF</b>     | 3.87             | <b>Moderate</b>                |
| <b>SMCAUSlsa</b> | 3.45             | <b>Moderate</b>                |
| <b>CRFAOa</b>    | 2.98             | <b>Low</b>                     |
| <b>CRFAOI</b>    | 2.67             | <b>Low</b>                     |
| <b>SMCAUSwn</b>  | 1.89             | <b>Acceptable</b>              |

The VIF analysis uncovers strong multicollinearity problems among the predictors. For instance, several features show severe multicollinearity ( $VIF > 10$ ), with PCDCz holding the highest VIF value of 188.51. High multicollinearity, with values from 5 to 10, is shown by the features of PCNARz (44.54), PCNARp (39.87), SMCAUSr (34.60), CNCCaus (31.99), and PCVERBz (30.94). Meanwhile, CNCLogic (23.72), PCREFz (17.04), and SMINTER (16.34) show moderate multicollinearity ( $VIF$  3-5). This high multicollinearity implies that many narrativity features are strongly correlated, leading to perturbation in coefficient

estimates, inflated standard errors, difficulty in interpreting individual predictor effects, and a reduction in the model's reliability.

## Residual Analysis



### Coherence & Cohesion (CC) Model - Residuals vs Fitted

The residuals roughly follow the expected pattern; i.e., they spread around the zero line with mild heteroscedasticity, as the variance appears to be slightly larger for higher fitted values (5.0-5.5). No strong curvature is visible; hence, it supports the assumption of a linear relationship. A few potential outliers are evident, but they are not extreme. The CC model yields acceptable residual patterns, albeit with some minor issues related to homoscedasticity, indicating that it is fairly well-specified among the four models.

### Grammar (GR) Model - Residuals vs Fitted

There is a slight non-linearity in nature, and the residuals appear to spread widely at higher values. The above abnormality makes the error variance non-constant, which can influence the model coefficients, and the resulting standard errors and confidence intervals can no longer be considered valid. To resolve such issues, one can try applying a variance-stabilizing transformation or robust regression methods.

### The Comparative Assessment

The CC model respects more assumptions of regression, while the GR model shows more serious diagnostic problems. Some heteroscedasticity is found in both models, but it is of more importance in the GR model. Also, the effect of outliers appears to be stronger in the GR model. The residual analysis confirms the previously highlighted validity issues of the multiple regression models, providing a significant advantage to Ridge regression for analyzing the relationships between narrativity features and IELTS writing assessment scores.

## Ridge Regression Results

### Model Performance

The table below presents the optimal regularization parameter ( $\alpha$ ), along with  $R^2$  scores and cross-validation results for each writing component.

| <i>Writing Descriptor</i>            | <i>Optimal Alpha</i> | <i>R<sup>2</sup> Score</i> | <i>Cross-Validation Score</i> | <i>Model Performance</i> |
|--------------------------------------|----------------------|----------------------------|-------------------------------|--------------------------|
| <i>Coherence &amp; Cohesion (CC)</i> | 10.0                 | 0.542                      | 0.485                         | <b>Best Performance</b>  |
| <i>Grammar (GR)</i>                  | 100.0                | 0.476                      | 0.421                         | Moderate Performance     |
| <i>Lexical Resource (LR)</i>         | 10.0                 | 0.530                      | 0.473                         | Strong Performance       |
| <i>Task Response (TR)</i>            | 10.0                 | 0.387                      | 0.329                         | Adequate Performance     |

The models accounted for between 38.7% and 54.2% of the variance in writing scores. The Coherence and Cohesion component has demonstrated the highest model performance, while Task Response has shown sufficient and lesser predictability.

### Ridge Regression Coefficients by Writing Descriptor

#### Coherence & Cohesion (CC) - $\alpha = 10.0$

| <i>Feature</i> | <i>Coefficient</i> | <i>Interpretation</i>                                                    |
|----------------|--------------------|--------------------------------------------------------------------------|
| <i>SMCAUSr</i> | 0.116              | Strongest positive predictor - causal cohesion ratio enhances coherence  |
| <i>CRFAOa</i>  | 0.089              | First-person pronoun usage positively impacts coherence                  |
| <i>PCDCz</i>   | 0.062              | Deep cohesion z-score contributes to coherence                           |
| <i>PCVERBz</i> | -0.164             | Strongest negative predictor - excessive verb cohesion reduces coherence |
| <i>SMINTER</i> | -0.045             | Intentional cohesion ratio negatively impacts coherence                  |
| <i>PCNARz</i>  | -0.035             | Narrative cohesion z-score shows negative effect                         |

Causal and deep cohesion bring forth a high degree of coherence, while repeated use of verbs and over-narrativization can weigh down fine idea development.

#### Grammar (GR) - $\alpha = 100.0$

| <i>Feature</i> | <i>Coefficient</i> | <i>Interpretation</i>                                   |
|----------------|--------------------|---------------------------------------------------------|
| <i>SMCAUSr</i> | 0.064              | Causal reasoning supports grammatical complexity        |
| <i>PCTEMPz</i> | 0.061              | Temporal cohesion enhances grammatical range            |
| <i>PCDCz</i>   | 0.038              | Deep cohesion contributes to grammatical sophistication |
| <i>PCVERBz</i> | -0.046             | Repetitive verb patterns reduce grammatical range       |
| <i>PCNARz</i>  | -0.032             | Excessive narrative focus may limit grammatical variety |

**Lexical Resource (LR) -  $\alpha = 10.0$** 

| <i>Feature</i> | <i>Coefficient</i> | <i>Interpretation</i>                                                        |
|----------------|--------------------|------------------------------------------------------------------------------|
| <b>PCVERBz</b> | -0.152             | Strongest negative predictor - repetitive verbs limit lexical sophistication |
| <b>PCNARz</b>  | -0.101             | High narrative cohesion may restrict lexical diversity                       |
| <b>SMCAUSr</b> | 0.103              | Causal reasoning enhances lexical resource                                   |
| <b>CRFAOa</b>  | 0.093              | Pronoun usage patterns positively affect lexical scores                      |
| <b>PCDCz</b>   | 0.071              | Deep cohesion supports lexical sophistication                                |

**Task Response (TR) -  $\alpha = 10.0$** 

| <i>Feature</i> | <i>Coefficient</i> | <i>Interpretation</i>                                                        |
|----------------|--------------------|------------------------------------------------------------------------------|
| <b>SMCAUSr</b> | 0.124              | Strongest positive predictor - causal reasoning crucial for task fulfillment |
| <b>PCREFz</b>  | 0.122              | Reference cohesion supports task response                                    |
| <b>PCNARz</b>  | -0.094             | Excessive narrative focus may detract from task requirements                 |
| <b>SMINTER</b> | -0.079             | Intentional cohesion patterns negatively impact task response                |
| <b>PCVERBz</b> | -0.065             | Repetitive verb usage reduces task response quality                          |

**Cross-Model Patterns:**

Some patterns hold consistently across writing components:

- **SMCAUSr (Causal reasoning):** The most consistently positive predictor, indicating that logical structure is fundamental to writing quality.
- **PCVERBz (Verb cohesion):** Strongly negative predictor under all the bands. Overuse of identical verbs may be a sign of a lack of lexical and syntactic variety.
- **PCNARz (Narrative cohesion):** Slight to moderately negative, hinting at the possibility that excessive narrative structuring in some way does not meet the expectations of the academic genre.
- **PCDCz (Deep cohesion):** Repeatedly had positive contributions, suggesting that writing-level discourse sophistication works in favor of writing quality.

**Model Validation and Stability:**

Regularization in the form of ridge regression addressed the multicollinearity issues and slightly improved model generalizability, with cross-validation scores increasing from 0.329 to 0.485, indicating a near absence of overfitting and good predictive stability of the models across writing components. Furthermore, the regularization shrunk unstable coefficients, thereby making models easier to interpret and enhancing robustness for further usage, such as automated scoring engines.

### Practical Implications:

The Ridge regression results strongly indicate that sophistication in causal reasoning (SMCAUSr) seems to be the most consistent predictor of writing quality. At the same time, repetitive verb usage (PCVERBz) appears to be a consistent negative predictor.

### Practical Implications for ESL Writing Instruction

The findings from the ridge regression models give explicit suggestions for teaching:

- **Emphasizing Causal Reasoning:** Instruct students in the development of cause-and-effect relations to provide greater clarity to argument and logical development.
- **Promoting Verb Diversity:** Discourage the overuse of common verbs; encourage the use of academic verbs (e.g., evaluate, argue, justify).
- **Balancing Narrative Elements:** Avoid over-narrating in academic writing; teach students to present ideas analytically rather than sequentially.
- **Developing Deep Cohesion:** Teach discourse markers and organizational devices (e.g., connectives, paragraph transitions) to raise overall cohesion.

Based on the above ridge regression analysis, it is inferred that causal reasoning and deep cohesion are the most reliable key indicators of high-quality academic writing. At the same time, repetitive verb use and excessive narrational focus significantly lower the quality of writing. These conclusions provide suggestions for enhancing ESL writing instruction and refining automated scoring systems to promote fairer, more linguistically aware assessment practices.

## 5. Discussion

The results of the study highlight the important role narrativity features, such as verb cohesion, narrative cohesion, and causal reasoning, play in predicting Pakistani ESL students' writing performance. Exploring the relationships between these linguistic features and IELTS-like descriptors for Coherence and Cohesion (CC), Lexical Resource (LR), Grammar (GR), and Task Response (TR), the study found clear-cut associations using Coh-Metrix indices.

The ridge regression models reported 38.7% to 54.2% variance explained in writing performance—an indication of model strength. Causal reasoning (SMCAUSr) was reported as the single strongest positive predictor across most writing components, while verb cohesion (PCVERBz) remained the most potent negative predictor. Narrative cohesion had somewhat inconsistent effects, supporting coherence but sometimes detracting from lexical or task-focused writing.

### Findings and Pedagogical Implications

#### Verb Cohesion and Writing Performance

The most consistent negative correlations between verb cohesion (PCVERBz) and band scores were observed for TR and LR, suggesting that students who tend to use repetitive or general-purpose verbs produce writing that lacks variety and precision. This dynamic directly complements recent findings where an elevated frequency of basic, highly common words significantly impairs Lexical Resource and Task Response scores (Bano et al., 2025), whereas expanding a text's semantic scope through deep semantic networks like hypernyms acts as a robust positive score predictor. Consequently, this supports prior research that has underscored lexical richness as a feature correlated positively with high-scoring writing.

To remedy this situation, the following activities are recommended for classroom use:

## 1. Verb Variety and Sentence Variation Practice

- **Verb Substitution Exercise:** This activity aims to diminish verb repetition and diversify language use for students. With lists of simple verbs given to them and encouraged to search for more specific, descriptive alternatives, students can develop their LR and hence aid in writing that is more varied and sophisticated.
- **Contextual Practice:** Short essays or paragraphs on a selected topic can be assigned to students, requiring them to use the same verb more than once. This makes them more conscious of the choice of verbs and how they contribute to fluent writing, resulting in better performance in the Lexical Resource section of the writing assessment.

## Causal Reasoning: The Strongest Positive Predictor

Causal reasoning features (SMCAUSr and SMCAUSlsa) associate most strongly and positively with writing quality, especially with task response and coherence and cohesion. Writers who properly use causal connectors tend to develop more logical, organized, and persuasive writing tools, which are vital for academic success. These activities help work on these skills:

## 2. Causal Reasoning Workshop

- **Cause-Effect Mapping:** The students can be presented with a series of events or scenarios and asked to map the cause-and-effect relationships that exist between them. This allows students to view the logical connections in their writing, thereby enhancing their ability to increase essay cohesion by providing clear, concrete arguments.
- **Essay Writing with Causal Chains:** From an academic standpoint, dealing with topics that employ cause-effect reasoning—such as "The impact of social media on education" or "How climate change is affecting agriculture"—gives students the opportunity to experiment with logical sequencing in their writing. This means that such an activity would help elevate Task Response (TR) by encouraging writers to ensure that their arguments are supported by evidence and that their ideas are expressed in a clear and logical sequence.
- **Peer Review on Causality:** Peer review sessions in which participants focus on how clear and strong the causal arguments are can significantly help students critically assess their application of causal reasoning, thereby fortifying their writing.
- **Narrative Cohesion: Useful but Must Be Balanced:** While narrative cohesion (PCNARz) had a beneficial effect on coherence and cohesion, the measure obtained a negative score for Task Response and Lexical Resource in the regression models. This reveals that students might rely too much on storytelling structures when writing academic-style essays—for instance, in IELTS essay writing—so much so that it detracts from the clarifying, objective, argumentative focus that style demands.

The following activity aims to help students find this balance:

## 3. Narrative Cohesion and Academic Writing

- **Narrative vs. Academic Writing Practice:** Students learn about cohesion by comparing narrative and academic essays on the same topic. They practice rewriting a narrative essay into an academic essay, emphasizing the reduction of superfluous storytelling elements and the use of formal devices to achieve academic cohesion.

- **Rewriting Exercise:** Students can be asked to rewrite them in academic style from simple narrative paragraphs. The exercise teaches the students to eliminate unnecessary elements of storytelling while keeping the logical flow of ideas. In doing so, they will learn how to ensure coherence in their arguments, a critical skill they must develop for any high-stakes test, such as the IELTS.

### Broader implications and challenges

These researches again confirmed the significance of specific language features in writing, and the students voiced some challenges that were peculiar to Pakistani ESL learners:

- **Repetitive verb overuse:** This tendency to over-rely on familiar verbal frames matches cognitive avoidance strategies documented locally, where writers purposefully downscale or simplify structural targets to manage external cognitive load and avoid making overt grammatical errors (Mahmood et al., 2026).
- Use of descriptive/narrative structure over argumentation
- Lack of acquaintance with advanced cohesive devices.

The challenge needs to be tackled by shifting the focus from mere grammar and vocabulary instruction to writing strategies at the discourse level, leveraging corpus-based tools such as Coh-Metrix. Once introduced, these tools could enable teachers to help students make concerted and measurable improvements in organizing, connecting, and logically developing ideas.

### 6. Conclusion

This paper has focused on how the features of narrativity, namely verb cohesion, narrative cohesion, and causal reasoning, influenced the writing performance of Pakistani ESL students. Findings exhibit that although verb cohesion (more specifically, verb repetition) has adverse implications on the quality of writing, causal reasoning has positive implications on coherence and cohesion (CC) and task response (TR). Although narrative cohesion is essential in structuring ideas, it should be combined with an academic style to ensure that one does not exaggerate it. The research lends confidence to the need for addressing these characteristics in ESL teaching. The use of pedagogical tasks, including verb substitution, cause-and-effect plot mapping, and the application of narrative to academic writing, will enable students to recognize the broader application of verbs, argument construction, and sentence-to-sentence coherence. The activities based on the corpus help enhance performance in writing tasks that resemble IELTS-like examinations.

### References

- Ahmad, M. (2022). Exploring disciplinary variation in Pakistani academic writing: A corpus-based research on doctoral dissertations. *Pakistan Languages and Humanities Review*, 6(IV). [https://doi.org/10.47205/plhr.2022\(6-iv\)06](https://doi.org/10.47205/plhr.2022(6-iv)06)
- Bano, N., Khan, M. M., Mahmood, R., & Mahmood, M. A. (2025). Exploring word-concreteness features as predictors of writing quality and their pedagogical implications. *Qualitative Research Journal for Social Studies*, 2(3), 944–962.
- Cheung, H. L. (2023). The role of narrative cohesion in children's narrative writing. *Journal of Writing Research*, 15(2), 121–136. <https://doi.org/10.1016/j.jwres.2022.12.005>
- Crossley, S. A., & McNamara, D. S. (2011). Predicting the writing quality of ESL learners using Coh-Metrix. *The Modern Language Journal*, 95(3), 371–387. <https://doi.org/10.1111/j.1540-4781.2011.01180.x>

- Gill, A. A., Kausar, G., & Haider, S. (2022). A corpus-based analysis of interpersonal persuasive writing skills of Pakistani English language learners. *Pakistan Journal of Social Research*, 4(2), 50–59. <https://doi.org/10.52567/pjsr.v4i2.451>
- Golparvar, S. E., Hu, G., & Seyedi, S. E. (2024). *Beyond the sentence: A cross-disciplinary investigation of cohesion in the discussion section of research articles*. SSRN. <https://doi.org/10.2139/ssrn.4825343>
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2004). Coh-Metrix: Providing multilevel analyses of text characteristics. *Discourse Processes*, 38(3), 247–280. [https://doi.org/10.1207/s15326950dp3803\\_2](https://doi.org/10.1207/s15326950dp3803_2)
- Haider, S. (2022). A corpus-based analysis of coherent writing skills of Pakistani English language learners. *Pakistan Languages and Humanities Review*, 6(II). [https://doi.org/10.47205/plhr.2022\(6-ii\)76](https://doi.org/10.47205/plhr.2022(6-ii)76)
- Lynch, T. (2001). *Study listening: A course in listening to lectures*. Cambridge University Press.
- Mahmood, M. A., Bano, N., Imran, M., & Almusharraf, N. (2026). Exploring the relationship between syntactic complexity, error types, and pedagogical implications in advanced writing: A quantitative analysis. *Reading & Writing*, 17(1), Article a587. <https://doi.org/10.4102/rw.v17i1.587>
- Ngo, T. T.-N., & Chen, H. H.-J. (2024). The effectiveness of corpus use in ESL/EFL writing: A meta-analysis. *Language Teaching Research*. Advance online publication. <https://doi.org/10.1177/13621688241260183>
- Qasim, H. M., Batool, A., & Nawaz, M. S. (n.d.). A corpus-based study of conjunctive cohesion in Pakistani research articles. *International Journal of Linguistics and Culture*. Advance online publication. <https://doi.org/10.52700/ijlc.v1i2.19>
- Ramzan, M., Mushtaq, A., & Ashraf, Z. (n.d.). Evacuation of difficulties and challenges for academic writing in ESL learning. *Journal of Development and Social Sciences*. Advance online publication. <https://doi.org/10.33195/maxskq26>
- Rauf, S., & Iqbal, H. M. (2008). Power of linguistic privilege: Critical discourse analysis of the narratives of Pakistani immigrant students in American schools. *Journal of International Education Research*. [http://pu.edu.pk/images/journal/ier/currentissue\\_pdf/4\\_Power%20of%20Linguistic%20Privilege.pdf](http://pu.edu.pk/images/journal/ier/currentissue_pdf/4_Power%20of%20Linguistic%20Privilege.pdf)
- Rupley, W. H., Kumar, A., McKeown, D., Syed, H., & Paige, D. D. (2023). *Beyond the red pen: Using dynamic assessment to mediate writing mechanics issues among Pakistani ESL learners*. SSRN. <https://doi.org/10.2139/ssrn.4566952>
- Sanosi, A. B. (2024). Revisiting cohesion in academic writing: A corpus-based analysis of EFL learners' use of conjunctions. *Indonesian Journal of Applied Linguistics*. Advance online publication. <https://doi.org/10.17509/ijal.v14i1.70418>
- Villarroel-Henríquez, V., & Bruna-Jofre, D. (2024). Levels of dialogicity and narrativity in teacher discourse and its relationship with student learning among high school students in a language class. *Cultura y Educación*, 36(1), 81–105. <https://doi.org/10.1177/11356405241235072>
- Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.
- Yasmin, T., Butt, I. H., & Sarwar, M. N. (2020). A comparative analysis of reporting verbs in research papers authored by Pakistani and native writers. *Global Linguistic Review*, 5(1), 57–66. [https://doi.org/10.31703/GLR.2020\(V-I\).07](https://doi.org/10.31703/GLR.2020(V-I).07)