

Spellings and Syntax in Crisis: A Corpus Based Study of ESL Challenges in Pakistani Middle Schools

Marryam Naz¹, Muhammad Farukh Arslan², Ahsan Mehmood Akbar³ and Rimsha Saeed⁴

<https://doi.org/10.62345/jads.2025.14.2.59>

Abstract

This study explores spelling accuracy among Pakistani ESL learners through a corpus-based analysis of spontaneous essay writing. One hundred and thirty students from grades 6, 7, and 8 were given an impromptu task to write a ten-line essay on the commonly taught topic: My Country: Pakistan. While the essay task is routinely introduced as early as grades 2 and 3 in Pakistani schools, a significant number of participants in this study exhibited notable spelling errors, including frequent misspellings of fundamental words such as “Pakistan” and the names of provinces. These morphological errors, rather than broader grammatical mistakes, formed the central focus of the investigation. Drawing on Error Analysis and Corpus Linguistics frameworks, this study constructed a small learner corpus from the collected essays and annotated the morphological mistakes for frequency, type, and pattern. The results reveal that spelling issues persist despite repeated exposure to the same content over several years, suggesting a gap between curriculum expectations and actual writing competence. The study raises questions about the effectiveness of school-level instructional methods, the absence of structured spelling pedagogy, and a possible overreliance on rote memorization. Findings also point to a lack of focus on form during writing instruction and minimal feedback strategies targeting spelling. This research contributes to both corpus-based error studies and educational policy critique, emphasizing the need for targeted spelling instruction and the integration of form-focused feedback in ESL classrooms across Pakistan.

Keywords: Corpus-based Spelling Analysis, Learner Corpora, Spontaneous Writing Corpus, Error-tagged Corpus, ESL Error Corpus, School-level Learner Corpus.

Introduction

Spelling, as a foundational element of written communication, plays a critical role in second language acquisition and literacy development. In the context of Pakistani education, however, spelling instruction is often overshadowed by grammar, vocabulary, and comprehension-focused practices. Despite its importance, morphological competence is rarely addressed as a distinct area of instruction, particularly during the early and middle stages of ESL learning. As demonstrated

¹MPhil (Linguistics) Scholar, Department of English, National University of Modern Languages, Islamabad (Faisalabad Campus). Email: nmasha13@gmail.com

²Lecturer, English Graduate Studies, National University of Modern Languages, Islamabad (Faisalabad Campus). Email: farukh.arslan@numl.edu.pk

³MPhil (Linguistics) Scholar, Department of English, National University of Modern Languages, Islamabad (Faisalabad Campus). Email: ahsanthabal@gmail.com

⁴BS English Linguistics, Department of English, National University of Modern Languages, Islamabad (Faisalabad Campus). Email: rimshasaeednuml@gmail.com



by Arslan et al. (2020) in their corpus-based study on derivational morphemes, Pakistani ESL learners exhibit a limited range and frequency of morphological structures compared to ENL speakers: highlighting systemic gaps in both instruction and learner acquisition. This underscores the pressing need to foreground spelling and morphology in educational practices, particularly at foundational stages.

This research explores spelling accuracy and pedagogical effectiveness by analyzing the spontaneous written responses of 130 students from grades 6, 7, and 8 on a universally familiar topic: My Country: Pakistan. This topic, frequently assigned as early as grade 2 in Pakistani schools, is considered both generic and culturally embedded and is often used to initiate essay writing skills among young learners. However, the present study reveals an alarming frequency of spelling errors, even in key lexical items such as the name Pakistan and its provinces: despite repeated exposure to the topic across multiple grade levels. This suggests a clear disconnection between exposure and acquisition: indicating that repetition alone does not guarantee morphological mastery.

The study is guided by theoretical frameworks from Error Analysis (Corder, 1967; James, 1998), Interlanguage Theory (Selinker, 1972), and Corpus Linguistics: focusing on the analysis of learner-generated texts to identify and categorize spelling error patterns. The task was intentionally spontaneous and unscripted: designed to assess learners' authentic written production without prior preparation or memorization. This allowed for genuine observation of learners' spelling behavior and the extent of their internalized morphological knowledge.

Methodologically, the study adopts both qualitative and corpus-annotated approaches to create an error-tagged learner corpus: categorizing mistakes by frequency, word type, and morphological pattern. The analysis not only quantifies the nature and recurrence of errors but also examines the cognitive and pedagogical factors that contribute to these deficiencies. It investigates how schools and teachers: through years of instructional engagement: have or have not prepared learners for such unsupervised, content-familiar writing tasks.

This study addresses a significant research gap in the Pakistani context, where little to no attention has been given to spelling errors in spontaneous middle school writing. While previous studies have addressed broader writing skills such as grammar, syntax, and vocabulary, the morphological (spelling) dimension remains underexplored: especially through the lens of learner-generated corpora. Moreover, there is a notable lack of research on student performance in unscripted writing scenarios on familiar topics: which better mirror real-world communicative needs.

By centering spelling as a core linguistic feature and employing a corpus-driven analysis, this study contributes meaningfully to both pedagogical reform and curriculum design. It highlights the urgent need for explicit, form-focused spelling instruction in Pakistani ESL classrooms: particularly in grades where foundational writing skills should be consolidating. Furthermore, it offers empirical insights for teacher training programs and curriculum developers: emphasizing the importance of aligning instructional practices with real-world writing expectations and learner realities.

Literature Review

The spelling and grammar issues found in spontaneous writing by Pakistani middle school ESL students reflect a convergence of well-established theoretical insights in second language acquisition. Spelling inconsistencies and phonetic approximations, particularly in high-frequency, culturally embedded words support early theories that emphasize the influence of first-language (L1) structures. Fries (1945) and Lado (1957) emphasized that learners subconsciously apply their

native language's structural patterns to English. Variants such as "Pakisatn," "Pokiston," or "Blochistan" suggest phonological transfer from Urdu and regional languages, reinforcing the notion that structural mismatches persist without adequate instructional mediation.

Corder's (1967, 1974) interlanguage model, later expanded by Selinker (1972), interprets errors as developmental stages rather than failures. Recurrent grammar issues, such as article misuse or fossilized verb forms illustrate how learners construct interim linguistic systems that require targeted feedback for progression. The lack of contextual writing activities and the dominance of rote learning contribute to the fossilization of these incorrect forms, aligning with Selinker's concept of interlanguage stagnation in the absence of corrective input.

The use of error analysis in identifying spelling patterns such as vowel omission, phoneme substitution, and consonant inversion validates earlier findings by Zughoul (1980), James (1998), and Al-Jarf (2010), who emphasized that error categories can serve as a roadmap for instructional improvement. Despite the pedagogical value of such analyses, the absence of classroom integration and teacher training in diagnostic error handling perpetuates recurring mistakes, particularly in under-resourced school systems.

Recent corpus-based research by Arslan et al. (2020) further reinforces the notion that linguistic competence among ESL learners remains limited in morphosyntactic development when compared to ENL counterparts. Their analysis of derivational morphemes across ENL, ESL, and EFL learners using ICNALE data revealed that ESL learners, specifically Pakistani participants exhibited reduced frequency and diversity in the use of grammatical word formations, particularly in noun and verb structures. This linguistic gap mirrors the syntactic and lexical challenges found in our student corpus, especially in the inconsistent or simplistic use of suffixes like -ment, -ness, or -ation, which are often absent or misused. The study's findings suggest that learners in ESL contexts lack the morphological richness necessary for complex syntactic construction, further limiting their writing fluency.

Cognitive approaches, including those of Cook (1991, 1997) and Gass and Selinker (2008), argue that spelling acquisition involves memory, cognitive load, and repeated exposure. Greater accuracy in spelling among students in print-rich environments or private institutions supports Nation's (2001) finding that word retention and spelling proficiency are strengthened by usage and frequent engagement. However, frequent exposure alone does not guarantee accuracy. Despite encountering key words like "Pakistan" across curricula, learners continue to misspell them, suggesting that repeated passive recognition is insufficient without deliberate practice and reinforcement.

Theories grounded in usage-based learning (Tomasello, 2003; Peters, 1985) support the idea that language is internalized through meaningful repetition. Observational data from schools indicate that many ESL writing activities lack context, creativity, or feedback, limiting opportunities for learners to experiment and internalize spelling and grammar patterns. This supports Dulay, Burt, and Krashen's (1982) emphasis on the importance of natural language acquisition environments, which are largely missing from teacher-centered, rule-heavy instructional models.

Orthographic irregularities in English, as emphasized by Crystal (2003), further complicate learning. When learners are presented with exceptions and variants without guided contextual exposure, spelling becomes more dependent on memorization than understanding. This condition is exacerbated by limited access to corrective feedback and insufficient teacher scaffolding, particularly in madrasa and government school systems.

Digitization efforts using OCR tools like Google Lens highlighted the lack of structural cohesion in student texts, with AI outputs often producing jumbled or incomplete transcriptions. This failure

demonstrates the depth of structural inconsistencies in learner writing and reflects broader issues of coherence and syntactic control. The AI's difficulty in processing learner texts indirectly supports Chater and Manning's (2014) conclusion that probabilistic, exposure-based models are more aligned with language acquisition than rule-driven frameworks, especially for orthographically complex languages like English.

The collective insights from foundational and contemporary linguistic studies converge with the observed error patterns in this corpus study. Persistent spelling and grammar errors are not random but are rooted in pedagogical, cognitive, and structural deficiencies. Without targeted interventions that draw on these theoretical foundations such as feedback-rich instruction, phonics-based training, and contextually grounded writing learners are unlikely to develop orthographic and grammatical fluency, despite repeated exposure in classroom settings. These findings, further supported by corpus-based evidence from Arslan et al. (2020), call for a reevaluation of instructional methods in Pakistani ESL education, aligning classroom practice with decades of SLA research.

Research Questions

1. What types of spelling errors do Pakistani ESL students in grades 6, 7, and 8 commonly commit in spontaneous writing tasks on the topic "My Country: Pakistan"?
2. Why do students continue to misspell familiar and frequently taught words such as Pakistan and the names of provinces, despite repeated exposure in school curricula from early grades?
3. Why have traditional classroom practices failed to ensure morphological level accuracy in unsupervised, real-time writing among middle school ESL learners in Pakistan?

Methodology

Step 1: Data Collection

The study involved a target group of 130 students drawn from three distinct school systems in Pakistan: government schools, madrassas, and local private institutions. The selected students were enrolled in grades 6 through 8, representing early middle school levels. A spontaneous writing task was administered, wherein each student was asked to compose a short essay on the familiar and culturally embedded topic, My Country: Pakistan. To ensure clarity, the task was orally explained for 5 to 10 minutes, outlining basic expectations and encouraging creativity.

Students were provided guiding prompts, allowing them to write about Pakistan's provinces, cities, cultural symbols, national identity, personal sentiments, or likes and dislikes. The goal was to assess spontaneous writing performance in an unprepared setting and capture authentic learner-generated language.

Step 2: Data Digitization

All student essays were handwritten on paper during classroom sessions. The digitization process involved using Google Lens, an OCR (Optical Character Recognition) tool, to convert handwritten text into editable Microsoft Word documents. Post-conversion, the outputs were compared against the original essays to address errors introduced by the scanning tool, such as jumbled phrases, distorted punctuation, and line breaks. These AI-induced errors were manually corrected to ensure that the digitized texts retained the students' original spelling and grammatical structures.

Step 3: Data Cleaning

Following digitization, a manual review was conducted to remove any personal identifiers, including names, dates, or irrelevant comments that could compromise student anonymity. In the

process of standardization, sentence boundaries and formatting were adjusted to correct any scanning inconsistencies. However, all spelling and grammatical errors made by students were preserved, as these were central to the analysis.

Step 4: Metadata Organization

Each cleaned and corrected essay was saved as a separate .txt file, with a systematic labeling convention indicating school type (e.g., GOV, MAD, PRV) and a unique student number. This metadata organization facilitated structured analysis, enabling school-wise comparisons and batch-level annotations.

Step 5: Corpus Compilation and Analysis

The compiled essays were imported into AntConc 4.2.0, a concordance and corpus analysis toolkit widely used in linguistic research. The resulting learner corpus comprised:

- Total Essays: 130
- Total Word Tokens: 15,530
- Total Word Types: 2,315

The following corpus-based analyses were performed:

- Frequency tagging of words and error forms
- Identification of recurring spelling and grammar mistakes
- Comparative analysis across school types to detect institutional error patterns

This five-step methodological framework ensured that the corpus was both systematically annotated and empirically grounded, allowing for reliable identification of trends in learner writing performance and morphological competence.

Analysis

Umul Madaris Corpus Section (Spelling and Grammar Errors)

The review of student essays from the Umul Madaris corpus, composed by middle-grade English language learners, reveals recurring patterns of spelling and grammatical errors, despite the familiarity and frequent instruction of the topic *My Country: Pakistan*. Although this topic is widely taught across grade levels and is culturally resonant, students continue to demonstrate substantial difficulties in accurately producing written English. This section presents a detailed analysis of the most frequently observed errors, categorized by type and frequency, while also exploring their instructional and cognitive implications. These findings are contextualized within established second language acquisition theories and supported by relevant literature, offering insight into the underlying causes of these persistent issues.

Grouped Spelling Mistakes

Several essential words related to national identity were consistently misspelled. This table summarizes some common mistakes:

Table 1: Common mistakes

Correct Word	Total Misspellings	Common Examples
Balochistan	13	blochistan, ballistan, balocihy
Pakistan	4	pakisatn, pokiston, paltistan
Sindh	3	sind, sinara, snidh
Khyber Pakhtunkhwa	3	khabarpaktunkhawa, kahberpactkawa, kpk
Country	3	contary, countary, contry

These error patterns suggest that learners encounter significant challenges in processing both phonological and morphological representations of English. As emphasized by Corder (1967) and James (1998), such errors are not merely accidental but reflect an underlying process in which learners construct interlanguage systems, interim grammatical frameworks influenced by both their first language (L1) and limited exposure to the target language (L2). These systems often generate systematic but non-standard forms as learners attempt to internalize rules and structures based on incomplete input.

For instance, in the Umul Madaris corpus, frequent misspellings such as *Pakisatn*, *Indpendet*, and *Blochistan* illustrate how learners rely on phonetic approximations rather than established morphological forms (see Table 1). These errors show consistent substitution patterns, often influenced by Urdu phonology and the absence of sound-letter correspondence instruction. The recurrence of such forms suggests that learners are drawing from an emergent rule set, one that prioritizes how words sound over how they are conventionally spelled in English.

This aligns with Selinker's (1972) theory of interlanguage development, which posits that learners will continue to produce such forms unless they receive targeted feedback and opportunities for restructuring. Without these, erroneous forms may become fossilized, especially in instructional environments that emphasize rote memorization over reflective writing practices. In this context, these misspellings are not signs of carelessness but rather developmental markers indicators of where the learner currently stands on the continuum of L2 acquisition.

Grouped Grammar Mistakes

Grammatical issues followed predictable categories. Below are some major types and examples:

Table 2: Major types of grammatical issues

Grammar Issue	Estimated occurrences	Examples
Missing Articles	9	"Pakistan is beautiful country", "flag is green and white"
Subject-Verb Agreement	6	"It have for provinces", "Pakistan are four province"
Verb Tense Errors	5	"Pakistan came into being 14 August", "It has came in 1947"
Capitalization Errors	5	"pakistan is my homeland", "i love my country"
Sentence Structure Issues	4	"I like very much I loveof my counsyt", "The is has many"

These errors suggest that learners are engaging in overgeneralization of grammatical rules and are experiencing interference from their native language (L1), most likely Urdu or a regional vernacular. Such tendencies are well-documented in second language acquisition research. According to Selinker's (1972) theory of interlanguage, learners construct provisional linguistic systems that are distinct from both their L1 and the target language (L2). These systems often reflect creative rule-making but are prone to instability and inconsistency, particularly in the absence of corrective feedback and contextualized input.

Examples from the Umul Madaris corpus, such as "Pakistan have four province", "It is the capital city is Islamabad", and "My country name Pakistan", demonstrate how learners overextend learned patterns or directly translate from L1 syntactic structures. These constructions reflect not only a

lack of syntactic fluency but also the influence of transfer errors, where the grammatical logic of the native language is imposed onto English. This phenomenon aligns with research by Zughouli (1980) and Odlin (1989), who noted that L1 influence is particularly pronounced when learners lack sustained exposure to accurate L2 input.

The fossilization of such errors is likely exacerbated by the inconsistency and superficiality of grammar instruction in many Pakistani classrooms. Grammar is often taught in isolation or through decontextualized rule drills, which fail to support the transition from receptive to productive competence. Without opportunities for meaningful practice, and in the absence of form-focused feedback, these interim structures remain unchallenged and unrefined, ultimately becoming entrenched in the learner's output.

Instructional Implications

The data reveals a clear dependence on repetition and rote memorization rather than communicative language practices within classroom instruction. While students can often recall the names of places, national symbols, and culturally significant terms, they consistently struggle to accurately represent these elements in written English. This disconnect between content familiarity and linguistic expression is particularly concerning, as it indicates that students are not internalizing the form-function connections needed for autonomous writing.

As Nation (2001) emphasizes, the acquisition of spelling and vocabulary requires reinforcement through varied, contextual use, not just exposure through copying or verbal recall. Similarly, Crystal (2003) notes that the irregularity of English orthography demands explicit instruction and guided practice; without these supports, learners are likely to reinforce incorrect forms through repetition. For example, in this study, commonly used words such as Pakistan, Balochistan, and independent were misspelled in multiple consistent but erroneous ways e.g., Pakisatn, Blochistan, and Indpendet, suggesting phonological overreliance and a lack of visual word recognition.

These findings are consistent with Al-Jarf's (2010) observation that EFL learners' spelling errors tend to follow predictable patterns, particularly when instruction lacks depth. Moreover, Cook (1999) views spelling as a developmental process that is influenced by both cognitive strategies and sociolinguistic environments. In this context, the high frequency of misspellings in words that are widely taught and culturally embedded points to significant deficiencies in early literacy instruction, especially in how spelling is introduced and reinforced during foundational years.

Curricular Relevance

While the content of student essays often reflects themes taught in the social studies curriculum, such as national leadership, cultural heritage, and sports, the linguistic expression of that content frequently falls short due to evident gaps in language training. Spelling errors such as "primonister" for Prime Minister and "hocley" for hockey illustrate a lack of phoneme-grapheme correspondence awareness, suggesting that students struggle to map sounds onto their correct written forms. These types of errors are indicative of insufficient phonological instruction and limited exposure to reading and writing tasks that reinforce correct spelling patterns.

The findings imply that students are able to recall and reproduce factual knowledge, a result of content-focused, memorization-heavy teaching methods but are not concurrently developing written language competence in English. This discrepancy supports the view of Mehmood and Rashid (2022), who argue that Pakistani ESL classrooms often emphasize rote memorization over integrative literacy development. Without structured opportunities for form-focused writing,

students fail to translate their subject knowledge into grammatically and orthographically accurate language.

Such patterns underscore the mismatch between content learning and language acquisition, emphasizing the need for interdisciplinary approaches that bridge curriculum subjects with targeted English instruction. Addressing this imbalance requires integrating language support strategies into subject teaching, ensuring that learners not only understand core concepts but also gain the linguistic tools to express them effectively in writing.

Analysis of Arqam School Corpus Section (Spelling and Grammar Issues)

This section analyzes essays written by Class 7 to 8, students from Arqam School on the topic of Pakistan. The findings highlight recurring spelling and grammar errors, with implications for curriculum development and classroom instruction.

Grouped Spelling Mistakes

Many critical words were misspelled in several forms, indicating that learners struggle with orthographic consistency, especially for proper nouns and culturally significant terms.

Table 3: Grouped spelling mistakes

Correct Word	Total Misspellings	Common Examples
Pakistan	22	pakis tan, Pakisti3n, Pakiston, pakstaon
Provinces	14	provincess, provines, provences, proviness
Resources	10	resourcess, resouces, iesousse
Quaid-e-Azam	10	Quaid-o-Azam, Quid-e-AZAN, Quadi-z-Azam
Balochistan	8	Balochistan, balochistan, Baleistan
Khyber Pakhtunkhwa	7	khyber paktoow khaw, Pakhtoonkham, khabar Pakhtokahan
Anthem	7	Antham, Aontham, annthm
Urdu	6	urduI, urdu., natral language urdu
Hafiz Jalandhri	6	hafiz jalndri, jalandari, haffiz jalandhri

These spelling inconsistencies reflect both phonological confusion and inadequate exposure to correct written forms. This aligns with Al-Jarf's (2010) claim that spelling patterns in EFL learners develop slowly and require explicit instruction and reinforcement through reading.

Grouped Grammar Mistakes

Grammar issues observed in this corpus are typical for early EFL writers and indicate developmental patterns:

Table 4: Grouped grammar mistakes

Grammar Issue	Estimated Occurrences	Common Examples
Subject-Verb Agreement	18	"It have five lovely provinces", "It flags white and green"
Article Misuse or Omission	15	"It is full resources", "The national poet is Dr. Allama"

Word Order Errors	13	"This kind is sweet culture", "Our flag colour in green and white"
Verb Tense Problems	12	"It come into existance", "Pakistan came existance"
Redundancy and Repetition	10	"It is it is full of resources", "Now we we are developing C pack too"
Capitalization Mistakes	8	"pakistan", "urdu", "allah almighty"
Misuse of Prepositions	6	"I live Pakistan", "gifted by good", "decorated with natural beauty by good"

These recurring errors are consistent with Selinker's (1972) theory of interlanguage, which posits that second language learners construct unique, transitional linguistic systems that are shaped by both their first language (L1) and the language learning environment. These interim systems often reflect internalized rules that diverge from both the target language (L2) and the learners' native language, resulting in predictable, systematic errors.

Within the corpus, students frequently demonstrate L1 interference, such as applying Urdu sentence structures or phonological patterns directly to English writing. For instance, constructions like "Pakistan have four province" or "I like Pakistan very much" reveal how learners map native syntactic logic onto English, producing forms that, while semantically intelligible, deviate from standard grammar. These examples reinforce Selinker's claim that interlanguage systems are not random but evolve based on learner hypotheses about language structure, which become fossilized when not explicitly corrected or restructured.

Such patterns underline the importance of providing learners with targeted, form-focused feedback, enabling them to revise faulty internal rules and progress toward greater L2 accuracy. Without deliberate pedagogical intervention, these interlanguage forms risk becoming entrenched, limiting long-term language development.

Educational Implications

Although the students demonstrate clear enthusiasm and content awareness, they face notable challenges in translating that knowledge into standard written English. The prevalence of repetitive sentence structures and limited lexical variation suggests an overreliance on memorized phrases, likely rooted in teacher-modeled constructions or textbook language. As Nation (2001) and Cook (1999) emphasize, learners require both enriched input and authentic output opportunities to build linguistic flexibility, syntactic range, and overall writing fluency. Without these, students default to familiar expressions, often at the expense of accuracy and creativity.

Furthermore, the Arqam School corpus reveals that learners frequently conflate religious, cultural, and national identities in their written expression. Words such as "gifted by Allah" and phrases like "IN SHA ALLAH" appear regularly in student texts, demonstrating the sociocultural embedding of English language use. As Crystal (2003) points out, such expressions are both culturally meaningful and pedagogically complex, posing unique challenges for ESL instruction. Teachers must navigate the tension between affirming cultural identity and fostering proficiency in standard written English.

Despite the familiarity of the topic, "My Country: Pakistan" students at Arqam School exhibited frequent spelling deviations and grammatical inconsistencies, indicating that even in private school contexts, literacy instruction may not sufficiently support orthographic or syntactic development. These patterns highlight the need for structured exposure to standard language forms, including

explicit instruction in grammar and vocabulary, particularly for transferring orally acquired knowledge into accurate written output.

This case-specific analysis provides a foundation for subsequent comparative evaluation across other school systems, particularly madrassa and government institutions. It also supports broader reflection on how instructional practices, rather than just curriculum content, directly shape learners' linguistic outcomes in ESL classrooms.

Analysis of Government School Corpus Section (Spelling and Grammar Issues)

This section presents a linguistic analysis of creative essays written by eighth-grade students from a government school in Pakistan. The data was collected from handwritten responses to the familiar classroom prompt "My Country: Pakistan". Despite the routine exposure to this topic within the national curriculum, the students' writing revealed frequent spelling and grammatical errors, indicating persistent challenges in written language production. The analysis identifies recurring error patterns, including misspellings of common nouns, inconsistent capitalization, and structural inaccuracies in sentence formation. These findings not only shed light on learner difficulties related to morphological and syntactic development but also raise important questions about the effectiveness of instructional practices in government school ESL contexts. The section further explores the pedagogical implications of these linguistic trends, suggesting the need for reform in how writing is taught, practiced, and assessed in such educational settings.

Grouped Spelling Mistakes

Spelling errors were systematic and repetitive, revealing core challenges in word recognition and production. Multiple incorrect forms of the same word suggest uncertainty in phoneme-grapheme correspondence.

Table 5: Grouped spelling mistakes

Correct Word	Total Misspellings	Common Misspellings
Pakistan	16	pakisatn, pakisatan, pokistan, patistan, paltistan
Province	12	provinces, prownse, provines, provincess
Beautiful	10	beautifull, buetyfull, beautiful, bouthyfull
Quaid-e-Azam	8	Quaide, Quid, Qaidi, Quaid-e-azem
Flag	7	flage, flagg, flagee
Urdu	5	urdu, urdo, urduI
Islamabad	5	islamad, islamaabad
Muslim	4	muslium, muslam, muslum
National	4	natainal, nataisla, natioal

The spelling variation observed among students in the government school corpus reflects a limited exposure to standardized print materials and a high degree of phonological confusion. Many students appear to rely on auditory recall rather than visual memory of word forms, resulting in inconsistent spellings of even frequently encountered terms. As Al-Jarf (2010) explains, such errors are largely rooted in the irregular nature of English orthography, which does not consistently reflect one-to-one sound-letter correspondence. Mastery of these irregularities requires extensive exposure to correct input, particularly through reading, and form-focused corrective feedback during writing tasks.

Within the corpus, examples such as “Patistan”, “Sinde”, and “provenness” illustrate how phonetic approximations replace standard spelling due to a lack of orthographic awareness. These deviations underscore a systemic instructional shortfall, wherein students are not regularly interacting with authentic reading materials or being guided through explicit spelling instruction. Without structured engagement with written language and opportunities for error analysis and correction, such spelling patterns are likely to persist and become fossilized.

Grouped Grammar Mistakes

The government school corpus reflects developmental issues typical of EFL learners, particularly in syntactic structure and verb usage.

Table 6: Grouped grammar mistakes

Grammar Issue	Estimated Occurrences	Common Examples
Subject-Verb Agreement	17	“Pakistan have many provinces”, “It have flag green and white”
Article Omission/Misuse	15	“Pakistan is national country”, “It is the capital city is Islamabad”
Word Order Problems	13	“My country name Pakistan”, “He is founder of Pakistan”
Verb Tense Confusion	12	“It come into being”, “He was make Pakistan”
Prepositional Errors	9	“I live Pakistan”, “gifted by good”, “go to the Lahore”
Capitalization Errors	8	“pakistan”, “allah almighty”, “urdu”
Redundancy and Repetition	6	“I love my my country”, “He he was leader”

The grammatical errors identified in the government school corpus clearly demonstrate the influence of first language (L1) transfer, most likely from Urdu or regional languages such as Punjabi or Sindhi. As theorized in Selinker’s (1972) interlanguage model, second language learners often construct interim linguistic systems that are shaped by both their native language and the learning environment. In this context, sentence structure and verb usage frequently reflect direct translations from Urdu, resulting in non-standard English constructions that are grammatically inaccurate.

For example, sentences like “It come into being in 1947” or “Capital city is Islamabad” reveal how learners apply native syntactic patterns and omit auxiliary verbs or articles, which are often optional or used differently in Urdu. These constructions are not random but follow predictable L1-influenced rules, reinforcing Selinker’s assertion that systematic interference plays a key role in shaping interlanguage forms. Unless learners are provided with explicit contrastive instruction and opportunities for error correction, these interlanguage structures may become fossilized, leading to persistent grammatical inaccuracies over time.

Such findings stress the importance of integrating contrastive analysis and corrective feedback into grammar instruction in Pakistani ESL classrooms, particularly in government schools where English input is often limited and largely rule-based.

Pedagogical Reflections

Despite the prevalence of spelling and grammatical errors, the student texts reveal a strong emotional connection to their homeland. Phrases such as “I love my country”, “Pakistan is my homeland”, and “I feel proud of Pakistan” frequently appear throughout the corpus, reflecting a deep sense of patriotism, identity, and conceptual engagement with the topic. However, students often lack the linguistic resources necessary to express these sentiments using grammatically correct English. This gap between emotional expression and linguistic accuracy highlights a fundamental challenge in ESL writing instruction, namely, connecting students' affective engagement with the structural demands of accurate written English.

The dominance of simple sentence constructions, repetitive vocabulary such as “beautiful”, “flag”, and “love”, and frequent direct translation attempts from Urdu highlight a pressing need for structured writing practice. As Nation (2001) emphasizes, effective writing development arises through iterative cycles of reading, imitation, and personalized production, wherein students are exposed to model texts and guided in replicating and adapting standard sentence structures.

The government school corpus exemplifies this tension between conceptual understanding and linguistic limitation. The recurrence of spelling and grammatical errors, many rooted in L1 transfer and insufficient morpheme-level knowledge suggests that learners are ready to express meaningful content but are constrained by weak language skills. Addressing these gaps requires integrated instructional strategies, including thematic vocabulary lists, targeted grammar drills, and exposure to authentic, well-constructed examples of student writing.

Ultimately, this corpus offers valuable insight for designing classroom interventions, particularly in under-resourced educational contexts where writing is often underemphasized. With targeted pedagogical support, government school students can move from conceptual fluency to linguistic competence, transforming spontaneous writing into a more accurate and expressive tool of communication.

Discussion

Table 7: School type number of spelling mistakes by students

School Type	Total Words	Unique Words	Total Spelling Mistakes	Frequent Spelling Issues	Examples
Madrassa (Umul Madaris)	4661	1003	312	Misspelling of 'Pakistan', province names, phonetic confusions	Pakisatn, Blochistan, Indpendet, countray
Government School	4893	1034	328	Misspelling of familiar terms, inconsistent capitalization	Patistan, Sinde, provencess, shopping
Private School (Arqam)	5976	1165	215	Mild phonetic misspellings, compound word errors	Pakistaan, Gilgite, urdo, shalwarqameez

Table 8: details of errors as type of school

School Type	Total Word Count	Total Grammar Errors	Subject-Verb Agreement Errors	Article Misuse	Tense Confusion	Preposition Errors	Word Order Errors	Example Errors
Madrassa	5021	96	26	22	18	8	12	"Pakistan have..." / "My country name Pakistan."
Govt	6088	145	41	38	29	12	15	"It come into being..." / "Capital city is Islamabad"
Private	4421	72	15	13	16	7	10	"Pakistan is a very beautiful" / "I like Pakistan very much"

This corpus-based investigation provides a comprehensive lens into the morphological and grammatical deficiencies prevalent in the spontaneous English writing of Pakistani middle school ESL learners. Drawing from a total of 130 essays across madrassa, government, and private schools, the study uncovers patterns of recurring spelling and syntactic errors that suggest significant instructional, cognitive, and systemic shortcomings. Through references to Tables 1 through 8, this discussion presents a theory-driven, comparative interpretation of the data, offering both diagnostic insight and strategic pedagogical recommendations.

The data in Table 7 highlight that government school students committed the highest number of spelling mistakes (328), followed closely by madrassa students (312), with private school students committing substantially fewer errors (215). More granular analysis in Tables 1, 3, and 5 indicates that even words central to national identity such as "Pakistan," "Balochistan," and "Quaid-e-Azam" were frequently misspelled. Variants like "Pakisatn," "Blochistan," and "Quid-e-AZAN" reveal not only issues of phonological confusion but also a lack of morphological familiarity with basic lexical items. These issues are especially evident in madrassa and government school corpora, where phoneme-grapheme mismatches dominate.

The prevalence of these errors in familiar vocabulary, particularly those that are repeatedly encountered in textbooks, prayers, and national celebrations, underscores a failure to bridge recognition and production. Baddeley's (2000) working memory theory explains that passive recognition is insufficient for encoding spelling into long-term memory unless paired with meaningful written production. The role of L1 interference is also critical; for example, learners transcribe English sounds like /sh/ or /zh/ using Urdu-based orthographic rules, leading to misspellings like "Balocihy" or "blochistan" (Table.1). Crystal (2003) stresses that English's irregular orthography exacerbates these challenges, especially when students lack explicit instruction in morpheme and phoneme interaction.

As Al-Jarf (2010) and James (1998) observe, EFL spelling errors are often predictable and systematic. The findings from Table.3 (Private school) and Table 5 (government school) reinforce this claim. For instance, the frequent distortion of "resources" into "resouces" or "iesousse" and the breakdown of compound terms like "Quaid-e-Azam" into fragmented and distorted segments indicate an insufficient grasp of word formation rules. Nation (2001) links this problem to a lack of morphological awareness and argues for the need to incorporate word family training and affix instruction in ESL curricula.

Morphological errors are not isolated from sociolinguistic context. In many essays from Arqam School (Table 3), terms like "Allah Almighty," "IN SHA ALLAH," and "natural language Urdu" occur frequently, reflecting students' cultural-linguistic landscape. While these expressions enrich learner identity, they also raise pedagogical challenges for spelling standardization and code-switching management, as noted by Canagarajah (2006) in his work on translanguaging.

In grammatical terms, Table 8 confirms that government school learners had the highest frequency of errors (145), followed by madrassa (96), and private schools (72). As shown in Tables 2, 4, and 6, common mistakes include faulty subject-verb agreement ("Pakistan have four province"), article misuse ("It is the capital city is Islamabad"), tense confusion ("It come into being"), and improper word order ("My country name Pakistan"). These align with Selinker's (1972) concept of interlanguage, where learners form evolving linguistic systems influenced by both their L1 and instructional environment.

Zughoul (1980) and Cook (1999) argue that the fossilization of incorrect grammatical forms often results from insufficient feedback and lack of production-based grammar instruction. The government school corpus (Table 6), which includes frequent repetitions and direct Urdu-to-English translations, supports this. For example, expressions like "He was make Pakistan" or "I love my my country" suggest surface-level grammar learning and translation dependence. Malik and Jabeen (2021) also emphasize that grammar rules taught in isolation do not result in mastery unless reinforced through authentic writing tasks.

Interestingly, while private school students performed better overall, Table 4 shows that they still struggled with verb tenses and article use. Errors like "Pakistan is a very beautiful" and "I like Pakistan very much" reflect awkward syntax and limited exposure to diverse sentence structures. This supports Tomasello's (2003) usage-based theory, which posits that learners acquire syntax through repeated use in meaningful contexts, a condition still underdeveloped even in better-resourced classrooms.

The comparative data across the tables underscore that instructional quality, rather than content exposure, determines learner performance. Students in all three systems showed a strong affective engagement with the topic, often expressing love for their country, its beauty, and history. Yet, as reflected in Tables 2, 4, and 6, the ability to express these sentiments in morphologically and syntactically accurate English was markedly inconsistent. Edge (1997) and Arslan et al. (2023) argue that affective familiarity with content must be coupled with structured language practice to produce effective writing.

To address these deficiencies, the study recommends the following:

- *Morphology-focused spelling instruction:* As seen in Tables 1, 3, and 5, frequent spelling errors in basic nouns and verbs require targeted instruction on roots, affixes, and sound-letter mappings.
- *Contextual grammar practice:* Grammar drills must evolve into writing-centered activities. Sentence restructuring, tense transformation tasks, and peer editing can bridge the theory-application gap evident in Tables 2, 4, and 6.
- *Learner corpus integration:* Teachers should use real student writing as a feedback tool. As suggested by James (1998), error analysis from learner corpora enhances metalinguistic awareness.
- *Curricular realignment:* At the policy level, writing fluency should be a core learning outcome. Formative assessment methods like error logs and writing portfolios must be embedded into the curriculum.

- *Teacher training:* Educators must be equipped with strategies to diagnose and respond to morphological and syntactic errors. Swan and Smith (2001) highlight the importance of understanding language transfer in shaping error correction.

Overall, the morphological and grammatical challenges identified in this study are not simply learner errors but indicators of structural gaps in ESL pedagogy. The persistent issues documented in Tables 1 through 8 demonstrate that passive exposure and rule memorization do not translate into productive competence. A reimagined ESL classroom—rooted in writing as process, form-focused feedback, and contextual instruction—is essential to bridge the gap between content familiarity and accurate expression.

It is important to recognize the efforts of all three educational sectors, particularly the madrassa and government school systems, which operate under substantial constraints yet demonstrated encouraging academic engagement. Although madrassas are primarily religious in focus, the institution included in this study, Umul Madaris, produced student essays that reflected patriotic enthusiasm and an emerging awareness of English writing conventions. These results suggest that with basic instructional scaffolding already in place, madrassas have the potential to further develop English literacy by integrating structured language instruction into their curriculum. Introducing systematic phonics-based spelling instruction, increasing student exposure to English print material, and providing teacher training in ESL pedagogy could significantly enhance learning outcomes in these settings.

Government school students, despite often facing limitations such as overcrowded classrooms, minimal access to reading resources, and insufficient teacher support, exhibited strong conceptual engagement and a genuine attempt to communicate ideas in English. Their writing reflected cognitive awareness but lacked structural precision. To improve outcomes in government schools, it is recommended that curriculum designers prioritize communicative writing activities over rote memorization, implement vocabulary-building strategies through contextual reading, and incorporate low-cost, high-impact writing tools such as learner-generated corpora and peer feedback practices.

The independent private school, Arqam, despite enjoying relatively better access to resources and teacher stability, still displayed gaps in learners' ability to apply grammar rules and spelling conventions in spontaneous writing tasks. This highlights that access alone is not enough. Teachers in such institutions should move beyond test-oriented instruction and encourage creative, process-based writing. Emphasizing formative feedback, expanding writing genres, and incorporating technology-assisted learning (such as digital dictionaries or writing games) may help strengthen orthographic and syntactic competence.

In sum, all three sectors demonstrate potential for growth, and with targeted interventions tailored to each school type's context and capacity, significant improvements in ESL writing proficiency can be achieved across Pakistan's diverse educational landscape.

Conclusion

The collective findings from this corpus-based investigation illuminate persistent and systemic shortcomings within Pakistan's ESL instructional landscape—particularly in the domains of morphological accuracy and syntactic control during spontaneous writing. Despite years of curricular exposure to familiar vocabulary and grammatical structures, students across madrassa, government, and private school systems continue to exhibit pronounced spelling and grammar errors, especially when writing in unscaffolded, real-time contexts.

These recurring mistakes are not isolated slips but rather the result of interwoven instructional, cognitive, and systemic limitations. Foremost among these is the widespread reliance on memorization-based pedagogies, which dominate madrasa and public-school classrooms. Instead of fostering critical engagement with English spelling systems or syntactic rules, instruction often defaults to rote copying and surface-level drills that leave learners ill-equipped for independent writing (Arslan et al, 2020). Compounding this issue is the near absence of individualized feedback, which stifles metacognitive development and prevents learners from recognizing, reflecting on, and correcting their errors (Malik & Jabeen, 2021).

From a linguistic standpoint, spelling errors such as “Pakisatn” or “Blochistan” stem from phonological misalignment and insufficient phonics instruction, as highlighted by Mehmood and Rashid (2022). On the grammatical front, problems like subject-verb disagreement, tense confusion, and misordered phrases reflect a gap between formally taught structures and practical application (Rehman & Mahmood, 2022). Further evidence from Arslan et al. (2020) shows that Pakistani ESL learners exhibit a notable deficiency in derivational morphology, producing far fewer nouns, verbs, and adjectives with affixation than their ENL counterparts—pointing to a deeper developmental lag in linguistic flexibility and fluency.

Crucially, these patterns expose a disconnect between language exposure and language acquisition. Learners may recognize familiar content orally or in recognition tasks, but lack the orthographic encoding and syntactic autonomy necessary to produce accurate, expressive written English. This confirms the theoretical assertions of Baddeley (2000) on working memory, Selinker (1972) on interlanguage fossilization, and on the importance of repeated, meaningful use for linguistic retention.

These findings underscore the urgent need for instructional reform that repositions writing as a central mode of learning, not just an outcome. Such reform should involve the incorporation of phonics-based spelling instruction, sentence-level grammar application, and corpus-informed feedback. Learners should engage in structured spelling routines, guided composition, peer revision, and reflective error analysis to internalize accurate patterns of usage.

More broadly, curriculum design must transition from passive knowledge delivery to productive language engagement, enabling learners to bridge the gap between oral familiarity and written precision. The ESL classroom must be reimaged as an active writing laboratory, a space where errors are not stigmatized but diagnosed, deconstructed, and corrected as part of a continuous developmental process.

Only through such a multidimensional, learner-centered, and feedback-rich model can Pakistani ESL students begin to achieve the morphological accuracy and grammatical fluency needed to thrive in academic and communicative domains.

References

- Abbas, A. A. (2010). Spelling errors among students of College of Education for Women. *Journal of the College of Education for Women*, 21(4), 855–866.
- Al-Ahdal, A. A. M. H. (2020). Error analysis of written English essays: The case of first year English majors at the University of Bisha, Saudi Arabia. *Arab World English Journal*, 11(4), 91–106.
- Al-Hamash, Kh. (1974). *English textbook in primary and secondary schools*. Baghdad: Institute for English Teaching, Iraq.
- Al-Jarf, R. (2010). Spelling error corpora in EFL. *Asian EFL Journal*, 12(2), 109–126.

- Al-Khaffaji, A., & Al-Shayib, M. (1987). *Analysis of English language primary textbooks in Iraq*. Baghdad: Ministry of Education Press.
- Almalki, A. (2021). The development of spelling skills in English as a foreign language: A longitudinal study. *Applied Linguistics*, 42(2), 195–218.
- Al-Zoubi, S. M. (2018). The impact of exposure to English language on language acquisition. *Journal of Applied Linguistics and Language Research*, 5(4), 151–162.
- Arslan, M. F., Mahmood, M. A., & Rasool, A. (2020, December). A corpus-based comparative study of derivational morphemes across ENL, ESL, EFL learners through ICNALE. *Linguistic Forum-A Journal of Linguistics*, 2(4), pp. 1-12).
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Bartram, M., & Walton, R. (2001). *Correction: A positive approach to language mistakes*. Heinle ELT.
- Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). Pearson Education.
- Carney, E. (1994). *A survey of English spelling*. Routledge.
- Chater, N., & Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in Cognitive Sciences*, 10(7), 335–344.
- Cook, V. J. (1991). L2 users and English spelling. *Journal of Multilingual and Multicultural Development*.
- Cook, V. J. (1999). *The English writing system*. Arnold.
- Corder, S. P. (1967). The significance of learners' errors. *International Review of Applied Linguistics*, 5(4), 161–170.
- Corder, S. P. (1981). *Error analysis and interlanguage*. Oxford University Press.
- Croft, C. (1983). *Teacher's manual for Spell-Write*. New Zealand Council for Educational Research.
- Crystal, D. (2003). *The Cambridge encyclopedia of the English language*. Cambridge University Press.
- Dulay, H., Burt, M., & Krashen, S. (1982). *Language two*. Oxford University Press.
- Edge, J. (1997). *Mistakes and correction*. Longman.
- Ellis, R. (1997). *SLA research and language teaching*. Oxford University Press.
- Fagerberg, I. (2006). *English spelling in Swedish secondary schools: Students' attitudes and performance*. Karlstad University Press.
- Hidalgo-McCabe, A., Manchón, R. M., & Gaviria, M. A. (2021). Spelling performance by Spanish primary students using English as a second language. *Frontiers in Psychology*.
- Hornby, A. S. (2000). *Oxford advanced learners' dictionary* (6th ed.). Oxford University Press.
- Invernizzi, M., & Hayes, L. (2004). *Word study in the classroom*. Pearson.
- James, C. (1998). *Errors in language learning and use: Exploring error analysis*. Longman.
- Lee, N. (1990). Notion of errors and appropriate corrective treatment. *Hong Kong Papers in Linguistics and Language Teaching*, 13, 1–15.
- Malik, F., & Jabeen, S. (2021). An analysis of common grammatical errors in English writing among Pakistani ESL learners. *International Journal of English Language and Linguistics Research*, 9(4), 23–35.

- Mehmood, M., & Rashid, A. (2022). Phonological misalignment and the neglect of phonics in Pakistani ESL classrooms: A critical review. *Journal of English Language Teaching and Linguistics*, 7(1), 45–59.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press.
- Rehman, S., & Mahmood, A. (2022). Grammar instruction and its practical disconnect in ESL writing: A study of learner errors in Pakistani classrooms. *Pakistan Journal of Language and Linguistics*, 6(2), 88–104.
- Peters, M. L. (1985). *Spelling: Caught or taught?* Routledge and Kegan Paul.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10(3), 209–231.
- Swan, M., & Smith, B. (2001). *Learner English: A teacher's guide to interference and other problems*. Cambridge University Press.
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Verhoeven, L., & Perfetti, C. A. (2021). Reading acquisition across languages and writing systems: An overview. *Journal of Neurolinguistics*, 59, 100992.
- Zughoul, M. R. (1980). Error analysis: A theoretical framework. *TESOL Quarterly*, 14(2), 204–210.