



A CORPUS-BASED COMPARATIVE ANALYSIS OF LEXICAL DIVERSITY OUTCOMES IN THE ENGLISH WRITING OF PUBLIC AND PRIVATE SECONDARY SCHOOL STUDENTS IN GUJRANWALA DISTRICT, PAKISTAN

*Syed Sajid Ali Shah¹, Muhammad Liaqat², Muhammad Tanveer Aslam³,
Choudhry Shahid⁴*

¹ PhD Scholar, Department of Linguistics, University of South Asia, Lahore, Pakistan. Email: sajidrw786@gmail.com

² PhD Scholar in English Linguistics, Department of English, University of South Asia, Lahore, Pakistan. Email: profliqat123@gmail.com

³ PhD English Linguistics, Visiting Lecturer, UOL Sargodha Campus, Pakistan
Email: tanveeraslamrana@gmail.com

⁴ Professor, Department of English, University of South Asia, Lahore, Pakistan
Email: shahid.mahmood@usa.edu.pk

Corresponding Author: Choudary Shahid

Email: shahid.mahmood@usa.edu.pk

Abstract

This study investigated the systemic disparities in productive English writing between public and private secondary school students in Gujranwala District, Pakistan, by empirically evaluating their lexical diversity outcomes. In Pakistan's stratified educational ecosystem, socio-economic divides drive structural imbalances between resource-constrained, public schools relying on the Grammar-Translation Method (GTM) and English-medium private schools utilizing communicative approaches. This study explores how these contrasting configurations shape active vocabulary deployment. The Gujranwala Secondary School Learner Corpus, which is a collection of the writings from students, was put together from a group of 100 secondary students. These students were in grades 9 and 10. The group was made to be fair and equal. It had 50 students from schools and 50 students from private schools. The group also had several boys and girls. The students had to write a story for 45 minutes. This was a test to see how well they could write. The Gujranwala Secondary School Learner Corpus is a useful tool. The Gujranwala Secondary School Learner Corpus has a lot of information about how students write. The linguistic analysis was grounded in Biber's Multidimensional Analysis Framework (Dimension 1: Involved vs. Informational Production) alongside Oxford's Strategic Self-Regulation (S2R) Model to evaluate vocabulary richness and communicative automaticity. Textual data logs were processed via AntConc (v4.2.4) and statistically verified using an Independent Samples t-test. The private school cohort demonstrated significantly greater lexical automaticity, generating a total volume of 11,450 running tokens ($\mu = 229.0$ words/essay) compared to the public cohort's 6,120 tokens ($\mu = 122.4$ words/essay). Surface lexical diversity measured via mean Type-Token Ratio (TTR) was significantly higher in private texts ($M = 0.61$, $SD = 0.05$) than in public texts ($M = 0.43$, $SD = 0.07$), $t(98) = 14.83$, $p < .001$, representing an exceptionally large institutional effect size (Cohen's $d = 2.96$). Private essays featured a higher part-of-speech content-word density (54.8%) compared to public sector essays (39.2%), which relied heavily on repeated structural function words (60.8%) and static, memorized examination templates. Private schooling yields substantially superior productive vocabulary richness, whereas public sector constraints encourage a restricted mental lexicon

and repetitive structural avoidance. Future research should pursue multi-district longitudinal lexical tracking using advanced natural language processing (NLP) architectures to monitor structural vocabulary development across regional domains.

Keywords: *Corpus Linguistics; Lexical Diversity; Type-Token Ratio (TTR); Biber's Multidimensional Framework; Stratified Educational Divide; ESL Writing in Pakistan.*

1. Introduction

Worldwide, English serves as a common way for people to communicate. Education relies on it heavily, just like work environments across fields. Schools in Pakistan require students to study English every year. This language usually guides daily lessons, mainly where fees are higher. Mastering how to express ideas clearly through writing matters deeply. A broad word choice strengthens that ability more than anything else. Knowing grammar helps shape sentences, yet choosing words matters just as much when putting thoughts on paper. How often someone uses different words shows how well they handle language overall. When a person sticks to the same few terms repeatedly, their range tends to be narrow. Using varied word options can signal stronger control over language structure. In Pakistan, how much learners learn English varies based on the kind of school they attend. Private institutions usually provide stronger support, fewer pupils per class, capable instructors, and greater chances to speak or read English every day. On the contrary, government-run schools often host bigger groups, limited supplies, and less frequent use of the language. As a result, young learners might develop uneven skills - some building wider vocabularies, others struggling more when it comes to putting thoughts into written form (Nation, 2022). Even though vocabulary matters a lot, few studies look at how public versus private school kids use words in Pakistan. Instead of tracking exams or grades, this work looks closely at real writings from high school students in Gujranwala. By comparing public and private schoolers through actual written samples, it reveals patterns not often seen before. The findings can quietly shape better ways to teach and help learners grow

1.1 Background of the Study

English language skills are really important these days for education, for getting a job and for talking to people around the world. In Pakistan, English is a language that people use in schools, government offices, and businesses. Students in schools and private schools learn English in different ways. This happens because of the way teachers teach English, how teachers are trained, and what schools think is important for English language skills. In Pakistan, people are worried because in government schools, English is taught in a way. Most of the time, English lessons are about grammar rules, translating things, memorizing words, and writing English. They do not focus on speaking English and listening to English, which is not good for English language skills. Also, English classrooms are usually controlled by the teacher, where students just sit and listen to the teacher, so they start to think of English as a subject they need to memorize or a language, they can use to get a job. As a result, many students have trouble talking to people in life using English, which is why this way of teaching English is not popular, as it does not help students learn English in a way they can actually use English. On the other hand, many private schools, especially in cities like Gujranwala District, use new and fun ways to teach English. They use things like talking to each other in English doing activities and using videos and pictures to help students learn English. That is why students in schools are usually better at English than students in public schools. Gujranwala is a place to look at this problem because it has private schools where students from diverse backgrounds study English. Old ways of teaching English are still used in some schools and new ways of teaching English are also used in schools. It is important to understand

how these old ways of teaching English affect how students learn English in schools compared to schools so we can find out what is missing and make education better, in the future for English language skills. English language skills are really important for students to learn. English is used everywhere in Pakistan, so students need to learn English in a way that they can actually use English to talk to people and get a job using English language skills.

1.2 Context of the Study

This study takes place in a district, namely Gujranwala, in Pakistan's Punjab province, which is noted for its rapidly rising educational industry and increasing need for English language ability, population is expected to be about 2,544,000 by 2025, representing a diverse socioeconomic, linguistic, and cultural community, district has a diverse selection of public and private secondary schools, both of which offer English as a required subject, the medium of instruction, method of instruction, and assessment standards contrast notably between the two educational sectors. Public schools in Gujranwala largely adhere to the regulations and curriculum guidelines established by the Punjab Education Department (PSED).

1.3 Rationale of the Study

Firstly, the discussion about the reality is, whether private or public schools provide better English language education has long been a significant issue for parents, teachers, and policymakers in Pakistan, rural and urban areas, such as Gujranwala. This issue becomes more relevant because people often spend large amounts of money on private schooling, based on the fact that private schools offer better education formally. So, these beliefs are often based on social perceptions rather than clear research evidence. Secondly, from an applied linguistics perspective, effective language learning depends on several approaches, such as teaching methods, classroom interaction, exposure to English, teacher competence, and students' socio-economic background. Because of these multiple factors, deciding which school system is "better" requires a careful comparison of both teaching practices and contextual conditions, rather than making general assumptions based on school type.

1.4 Statement of the Problem

There is a difference in English language skills between private and public schools in Pakistan. This is an issue in education. It is not about money. Many things contribute to this problem. It affects students' ability to learn English properly. We need to look at things like fairness, standards, and availability of resources. Private schools have a lot of money because they charge fees, and many people want to send their kids there. This means they can provide buildings, qualified teachers, and new teaching materials. So, they can give English education. On the other hand, public schools often do not have enough money. This leads to many kids in one classroom, old teaching methods, teachers who are not well-trained, and not enough materials to help kids learn. These things make English language teaching not good. Kids from families have it even tougher. Kids in schools usually have families who know a lot about education and speak English at home. This helps them learn better. Public school kids do not have these advantages. They do not get to learn in an environment. Private schools teach kids how to communicate in English and use methods. Public schools focus more on memorizing rules and grammar. This does not help kids use English in life. Also, parents of private school kids are more involved in their kids' English education. They want their kids to do well. They help them. English language skills are very important. Private and public schools are very different when it comes to teaching English.

1.5 Research Gap

There is a lot of research on how people learn English in Pakistan. We still do not know much about how private schools and public schools compare when it comes to English language education. Most studies look at either the quality of education or the social and economic factors that affect learning. They do not look at everything, like teaching methods, school buildings, and the social and economic backgrounds of the students in private schools and public schools. For example, some studies have found that private schools are better at teaching English. We do not know exactly how things like what is taught, how good the teachers are, and how much parents are involved affect the differences between private schools and public schools in Gujranwala District. We need studies that look at how all these things work to affect how well students learn English, especially in secondary school. Now, most studies just use numbers. Do not ask students, teachers, and parents what they think, which means we do not really understand the problems that both private schools and public schools face. This makes it hard for people who make decisions about education to come up with plans to help students in schools and students in private schools learn English. English language education is very important. This study wants to fill in the gaps by comparing how private schools and public schools in Gujranwala teach English using both numbers and stories from students, teachers, and parents to get a picture of why some schools are better at teaching English than others. The study will look at language learning outcomes in schools and public schools to understand what causes the differences. By doing this, the study will provide an understanding of English language education in Gujranwala District.

1.6 Purpose of the Study

The purpose of this study is to compare the English learners' results in public and private secondary schools in Gujranwala, Pakistan. It checks student results, teaching methods, family backgrounds, and other resources. This study wants to know what causes the skill gap. It looks at school factors, social background, and the teaching styles that affect learning English. The goal is to help teachers, school managers, leaders, and decision-makers understand these differences. They want to improve education fairly in both private and public schools. The study aims to give a picture of English learning in private and public schools. It looks at what causes differences in English skills. The goal is to provide help to make English education better. Finally, the study compares English learning results in public schools. It wants to help teachers and leaders improve the best strategies in the education sector.

1.7 Research Objectives

The main focus of this research is to compare the English language competency of students studying in private and public secondary schools in Gujranwala. It carefully examines how well students can read, write, use vocabulary, and understand English in writing structure in both types of schools. This study also intends to look at the teaching styles and classroom methods used by teachers and identify which approaches are more helpful for improving students' learning. In addition, it will consider how school facilities such as libraries, technology, and other learning resources influence students' progress in English.

Another important purpose of this study is to explore how students' use of writing and its functions conditions affect their English learning. It will focus on how limited income and fewer educational opportunities may create challenges, especially for students in public schools. The research will also look at the role of parents in supporting their children's education. Based on the findings, the study will suggest practical and realistic recommendations for teachers and policymakers to help

reduce the gap in English proficiency between private and public schools and improve the overall quality of English education.

1.8 Research Questions / Hypotheses

Research Questions

1. What is the difference in lexical diversity in the English writing of public and private school students?
2. Which group of students uses more content-rich vocabulary in their writing?
3. What are the most commonly used words by public and private school students?
4. Is the difference in lexical diversity between public and private school students statistically significant?
5. How can the findings help improve the English writing skills of students?

Hypotheses

H1: Students in private secondary schools exhibit higher lexical diversity in English writing than students in public secondary schools.

H2: The mean Type-Token Ratio (TTR) of private school students is significantly higher than that of public-school students.

H3: Private school students use a higher proportion of content words compared to function words than public school students.

H4: Word frequency patterns show that private school students use a wider range of vocabulary, whereas public school students rely more on repeated common words.

2. Literature Review

Recent advancements in automated text analysis highlight that higher lexical diversity directly correlates with superior text quality, essay scores, and structural complexity (Crossley, 2020). When learners deploy a wider array of low-frequency content words (nouns, lexical verbs, adjectives, and adverbs), they display an advanced capacity to express precise, nuanced ideas without relying on repetitive, generic vocabulary. Conversely, text corpora characterized by low lexical diversity often reflect restrictive vocabulary instruction, structural avoidance strategies, and a reliance on formulaic language (Egbert et al., 2022). To systematically evaluate the variations in student writing across public and private educational sectors, this study relies on Biber's Multidimensional Analysis (MDA) framework adapted for learner corpora alongside modern multi-metric models of lexical complexity. Biber's model operates on the principle that linguistic features co-occur in predictable patterns, which reflect underlying functional, situational, and cognitive dimensions of language use (Biber & Egbert, 2024). Within this framework, written texts are evaluated across key dimensions: Involved vs. Informational Production: This dimension contrasts high-density informational discourse (marked by high lexical diversity, frequent nouns, and attributive adjectives) with involved, interactive discourse (marked by repetitive function words, pronouns, and verbs). Narrative vs. Non-Narrative Concerns: This dimension tracks the deployment of past-tense verbs and aspect markers to establish descriptive timelines. By integrating Biber's multidimensional perspectives with automated measures of lexical richness—specifically the Type-Token Ratio (TTR), the Root Type-Token Ratio (RTTR), and the Moving-Average Type-Token Ratio (MATTR)—this study avoids the computational biases introduced by varying text lengths (Kyle et al., 2021). This combined theoretical approach ensures that the

evaluation goes beyond surface-level spelling or mechanics. Instead, it measures the exact cognitive and functional depth of the vocabulary that secondary school students use when generating academic text. The educational infrastructure of Pakistan is deeply fragmented, creating distinct linguistic outcomes based on institutional design. The divide between public sector schools and private institutions represents not just a difference in funding, but two fundamentally distinct educational philosophies and pedagogical approaches toward English Second Language (ESL) acquisition (Ahmed & Malik, 2021). In the vast majority of public secondary schools across Punjab, English instruction continues to be dominated by traditional, translation-heavy methodologies. As documented by Khan (2020), the Grammar-Translation Method is what most teachers use. They teach students the rules of grammar in class. Teachers give lectures in the students' own language, which's usually Urdu or Punjabi. The Grammar-Translation Method is about using English to practice translating sentences. Learning words is not the thing in the Grammar-Translation Method. Students have to memorize lists of words in both languages for tests. The tests are easy to prepare for because they are very predictable. Students just have to recall what they have learned by heart for the tests. The Grammar-Translation Method is still what most teachers use.

The Grammar-Translation Method is a way of teaching. (Saleem & Javed, 2023). This environment lacks opportunities for productive output, meaningful communication, or creative text generation, which limits students' ability to internalize and use an active, diverse vocabulary. Conversely, private secondary institutions generally operate with greater curricular flexibility, emphasizing communicative capability and interactive language immersion. Modern private schools actively implement elements of Communicative Language Teaching (CLT) and Task-Based Language Teaching (TBLT) (Ali et al., 2021). In these settings, English is utilized as the primary medium of instruction across all subjects, maximizing students' daily linguistic exposure. Vocabulary is taught contextually through intensive reading materials, collaborative presentations, and formative writing tasks. This immersive exposure encourages private school students to develop an active mental lexicon, allowing them to retrieve and deploy low-frequency words and descriptive adjectives more freely during writing assessments (Hussain, 2022).

Beyond classroom instruction, a learner's capacity to build and use a diverse vocabulary is closely linked to socio-economic realities and institutional resources. Recent sociological and educational literature in South Asia demonstrates that institutional stratification mirrors broader socio-economic divides (Raza & Shah, 2024). The physical and digital infrastructure available within schools plays a major role in shaping literacy outcomes. Private institutions often leverage higher tuition fees to provide student-centered resources, including modern libraries, digital multimedia tools, language learning software, and consistent formative assessment protocols (Umar & Mahmood, 2023). These tools provide students with regular exposure to varied, authentic English text corpora. In contrast, public schools frequently grapple with severe infrastructural challenges, including overcrowded classrooms where teacher-to-student ratios often exceed 1:60, an absolute lack of supplementary reading materials, and a complete absence of digital educational tools (Ahmed & Malik, 2021). In an over-crowded environment, teachers are forced to rely on choral repetition and standardized textbook scripts, which restricts individual linguistic experimentation and limits vocabulary growth. The home environment adds another layer to this educational disparity. Students enrolled in private schools predominantly come from middle- to upper-class backgrounds where parents are highly literate and often fluent in English. As noted by Bashir et

al. (2025), these parents actively engage in their children's academic lives, provide supplementary tutoring, purchase non-curricular literature, and foster a home environment where English carries significant cultural and social value.

Conversely, public school cohorts often draw from working-class or rural backgrounds where parents may have limited formal education and low English literacy. Consequently, these students receive minimal academic support at home and rely entirely on the school environment for language input. This lack of external reinforcement deepens the lexical deficit, as these learners rarely encounter opportunities to practice or expand their vocabulary outside of the school gate. Over the past six years, corpus linguistics has transformed the empirical evaluation of learner language by replacing subjective evaluations with precise, quantitative metrics. Globally, learner corpus research (LCR) has shown that non-native writers exhibit unique patterns of lexical over- and under-use compared to native benchmarks, and that these patterns are heavily influenced by instructional input (Flowerdew & Mahlberg, 2021). In the Pakistani context, recent corpus-based studies confirm a significant performance gap between different educational tracks. For example, a corpus analysis conducted by Saleem and Javed (2023) revealed that public school graduates frequently rely on high-frequency, generic function words (e.g., *good, bad, happy, do, go*) and produce texts with low Type-Token Ratios. Furthermore, contemporary research indicates that public sector writing displays a high degree of word repetition and formulaic structural framing, which points to a reliance on memorized essay templates designed for examinations (Raza & Shah, 2024). In contrast, private school corpora display higher lexical variation, greater lexical density, and a more balanced distribution of nouns, adjectives, and descriptive adverbs. This empirical evidence underlines the need for a targeted, localized corpus evaluation within Gujranwala District to understand how regional institutional structures directly impact students' productive vocabulary. Corpus-based studies have become an important tool in analyzing second language learning, allowing researchers to examine actual student writing rather than relying solely on test scores (Biber, Conrad & Reppen, 2011). Such studies help identify patterns in vocabulary use, sentence structure, and overall language proficiency. Lexical diversity is a widely used measure in language research, often calculated using Type-Token Ratio (TTR) or other indices, to indicate the richness of vocabulary in learner writing (Nation & Webb, 2011). High lexical diversity suggests a broad vocabulary, while low diversity points to repetition and limited word use. Several studies have compared English outcomes in public and private schools. Ali & Shah (2015) and Hussain (2018) report that private school students often outperform public school students due to better resources, smaller classes, and qualified teachers.

2.1 Corpus-Based Studies in Second Language Learning

Corpus-based studies use large collections of authentic texts, called corpora, to analyze how learners use a second language (Granger, 2002). These studies help researchers examine vocabulary, grammar, and writing patterns in real learner data rather than relying only on tests or classroom observations. Learner corpora allow comparisons across groups, showing how lexical diversity and language use develop over time. Research indicates that students with more exposure to English tend to use a wider range of vocabulary and more complex structures (Hunston, 2002). In Pakistan, corpus-based studies are limited, especially for secondary school students. Applying corpus analysis can reveal patterns in vocabulary use, compare public and private school learners, and provide insights for improving English teaching (Granger, 2002; Hunston, 2002).

2.2 Lexical Diversity and Its Measures

Lexical diversity refers to the variety of words a person uses in writing or speaking. It is an important indicator of vocabulary knowledge and language proficiency. A text with high lexical diversity shows that the writer can use many different words, while low diversity suggests limited vocabulary (McCarthy & Jarvis, 2010). There are several ways to measure lexical diversity. The most common is the Type-Token Ratio (TTR), which divides the number of different words (types) by the total number of words (tokens) in a text. A higher TTR indicates more diverse vocabulary (Malvern & Richards, 1997). Other measures include MAT, which adjusts for text length and gives more reliable results for longer texts. Lexical diversity has been widely used in second language research to assess learners' writing skills, compare groups, and track progress over time. For example, learners with more exposure to English tend to produce texts with higher lexical diversity, reflecting richer vocabulary knowledge (Crossley et al., 2011). Using these measures allows researchers to objectively evaluate vocabulary use, identify strengths and weaknesses in students' writing, and compare learners from different educational backgrounds, such as public and private schools.

2.3 Studies on Public vs Private School English Outcomes

Several studies have highlighted differences in English learning outcomes between public and private school students. Research shows that private school students often perform better in English, particularly in vocabulary, reading, and writing, due to better resources, smaller class sizes, and more exposure to English both inside and outside the classroom (Ali & Shah, 2015). Public school students, however, may face challenges such as limited teaching materials, large class sizes, and fewer opportunities to practice English, which can affect their writing skills and overall language proficiency (Hussain, 2018). These differences are particularly visible in lexical diversity, grammar accuracy, and fluency in writing tasks. Some studies have used standardized tests, while others have examined student essays to identify differences in vocabulary use and sentence structure. Findings suggest that school type significantly influences English outcomes, but factors such as teacher training, teaching methods, and student motivation also play a role (Khan, 2019). Understanding these differences is important for improving teaching strategies and providing targeted support for students, especially in public schools where resources are limited. Corpus-based analysis can offer a more detailed view of how students actually use English rather than just relying on exam scores.

2.4 Learner Corpora and AntConc Applications

Learner corpora are collections of texts produced by language learners, often used to study patterns in vocabulary, grammar, and writing skills. These corpora help researchers understand how learners develop language over time, identify common errors, and compare performance across groups (Granger, 2003). One popular tool for analyzing learner corpora is AntConc, a free concordance program that allows researchers to examine word frequency, collocations, keywords, and lexical diversity. AntConc is widely used in applied linguistics because it is user-friendly and effective for analyzing large amounts of text (Anthony, 2020). For example, researchers can use AntConc to measure Type-Token Ratio (TTR), identify the most frequently used words, and compare lexical patterns between students from different schools or proficiency levels. This makes it possible to objectively evaluate vocabulary use and writing quality. In the context of Pakistan, AntConc can be particularly useful for comparing public and private school students, providing insights into how school type affects English writing. By systematically analyzing student essays,

researchers can identify strengths, weaknesses, and patterns of language use that are not visible through traditional exams.

2.5 Gaps in Existing Research

While there is substantial research on second language learning and corpus-based studies globally, limited research exists in Pakistan focusing on the lexical diversity of secondary school students' English writing. Most studies compare exam scores or general proficiency rather than analyzing how students actually use language in writing (Ali & Shah, 2015; Hussain, 2018). Moreover, few studies have directly compared public and private school learners using corpus-based methods. This means there is a lack of detailed evidence on vocabulary richness, word choice patterns, and differences in writing style between these groups. Additionally, while tools like AntConc have been widely used internationally, their application to Pakistani secondary school student corpora is rare. These limits understanding of how teaching methods, school environment, and exposure to English affect students' writing skills. This gap highlights the need for a corpus-based study that systematically analyzes student essays, compares public and private school learners, and provides insights for improving English teaching and learning in Pakistan.

3. RESEARCH METHODOLOGY

3.1 Research Design

This study adopts a quantitative, corpus-driven comparative research design to examine the lexical diversity outcomes in the English writing of public and private secondary school students. A corpus-based methodology allows for an objective, empirical analysis of authentic language production, moving past the limitations of traditional impressionistic grading (Egbert et al., 2022). By compiling and analyzing specialized learner corpora, this research systematically measures and compares explicit structural indices of vocabulary deployment across two distinct educational tracks.

3.2 Target Population and Sampling Matrix

The target population comprises secondary school students (Grades 9 and 10) enrolled in public and private institutions within District Gujranwala, Punjab, Pakistan. Sampling Strategy: A stratified purposive sampling technique was used to ensure an equal distribution of institutional parameters, geographic representation, and gender dynamics. Sample Size (N): The study sampled a total of N = 100 secondary school students. Strata Distribution: The sample was divided evenly into two institutional groups: Public School Cohort (n = 50): Selected from representative institutions regulated by the Punjab School Education Department (PSED).

Private School Cohort (n = 50): Selected from decentralized English-medium private institutions operating in the same urban and semi-urban sectors.

Table 3.3: Sampling and Participant Demographics

Institutional Sector	Grade Level	Male Participants	Female Participants	Total Essays Collected
Public Sector	Grades 9 & 10	25	25	50 Essays
Private Sector	Grades 9 & 10	25	25	50 Essays
Total Sample	—	50	50	100 Essays

3.4 Corpus Design and Compilation

To build a reliable empirical database, a specialized learner corpus—the Gujranwala Secondary School Learner Corpus (GSSLC)—was constructed. The corpus compilation followed standard international guidelines for learner corpus design (Flowerdew & Mahlberg, 2021). Elicitation Task: Participants were required to produce a descriptive or argumentative essay on a standardized topic ("*My School*" or "*Life in a Pakistani City*"). This controlled prompt minimized topic-induced lexical variation. Operational Controls: Time allocation was strictly capped at 45 minutes. The use of dictionaries, mobile phones, textbook references, or instructional assistance was strictly prohibited to guarantee that the output reflected the students' active mental lexicons. Corpus Processing: The handwritten essays were transcribed verbatim into plain text files (.txt) encoding UTF-8 format. Crucially, original spelling and grammatical anomalies were preserved during transcription to avoid artificially altering the lexical indices.

3.5 Instrumentation and Analytical Software

The completed digital corpus was processed using AntConc (Version 4.2.4 or later), a widely recognized software program for corpus analysis (Anthony, 2024). This tool allowed for the automated extraction of exact word counts, frequency lists, and key linguistic parameters. The analytical focus centered on two operational dimensions derived from Biber's Multidimensional Framework and modern computational linguistics (Biber & Egbert, 2024; Kyle et al., 2021)

Type-Token Ratio (TTR): Calculated as the number of unique words (Types) divided by the total number of running words (Tokens). This serves as the primary mathematical measure of surface lexical diversity:
$$\text{TTR} = \frac{\text{Total Unique Types}}{\text{Total Running Tokens}}$$

Lexical Density: Measured by calculating the precise ratio of content words (nouns, main verbs, adjectives, and adverbs) relative to structural function words (articles, prepositions, and conjunctions).

3.6 Data Analysis Procedures

Once processed through AntConc, the quantitative data logs were exported into statistical analysis software. Descriptive statistics (means, standard deviations, maximum and minimum ranges) were compiled for both school sectors. To determine whether the observed differences in lexical diversity between public and private school student texts were statistically meaningful, an Independent Samples t-test was conducted. The alpha significance level was set at $p < .05$. This statistical confirmation ensures that the conclusions are mathematically verified and free from sampling bias (Raza & Shah, 2024).

4. DATA ANALYSIS AND RESULTS

The quantitative findings of this study are drawn from a computational evaluation of the *Gujranwala Secondary School Learner Corpus (GSSLC)*, which comprises 100 standardized descriptive and argumentative essays balanced evenly between public sector ($n = 50$) and private sector ($n = 50$) secondary school students. The corpus structural indicators, surface lexical diversity metrics, and part-of-speech distributions were extracted using AntConc (Anthony, 2024) and statistically verified to evaluate the operational hypotheses (\mathbf{H}_1 through \mathbf{H}_4).

4.1 Corpus Overview and Surface Descriptive Statistics

Initial structural processing of the GSSLC shows a notable difference in the total textual volume (Tokens) generated by the two student cohorts within the identical 45-minute testing window. The

4.5 Part of Speech (POS) Profile and Lexical Density

To answer RQ2 and test \mathbf{H}_3 , the texts were analyzed to determine the balance between content words (nouns, action verbs, attributive adjectives, adverbs) and functional elements (prepositions, conjunctions, determiners). This tracking aligns with Biber's structural dimensions, where informational clarity relies heavily on high content-word density (Biber & Egbert, 2024).

Private School Writing: Displayed an overall lexical density of 54.8%. These texts regularly feature rich descriptive phrases (e.g., "well-equipped laboratories", "spacious courtyard", "highly qualified educators"), signaling an active vocabulary.

Public School Writing: Displayed a lower lexical density of 39.2%. These texts rely more heavily on functional framing and static linking verbs (e.g., "is", "has", "have"), reducing the overall descriptive quality.

Table 4.6: Part-of-Speech and Density Breakdown

Grammatical Category	Private Sector Frequency (%)	Public Sector Frequency (%)	Category Functional Orientation
Nouns (Substantives)	24.5%	18.2%	Content / Informational
Lexical Verbs (Action)	14.2%	10.5%	Content / Narrative
Adjectives (Attributive)	10.6%	6.4%	Content / Qualitative Elaboration
Adverbs (Modifier)	5.5%	4.1%	Content / Circumstantial
Function Elements	45.2%	60.8%	Structural / Grammatical Framing
Total Target Focus	100.0%	100.0%	Lexical Density: Private (54.8%) vs. Public (39.2%)

4.7 Word Frequency Profiles and Repetition Patterns

To address RQ3 and evaluate \mathbf{H}_4 , top frequency lists were extracted using AntConc. This tracking illustrates that public sector texts feature a narrow range of vocabulary, relying heavily on a limited set of repeated words.

Public Sector Patterns: The raw frequency logs show extreme repetition of introductory structural forms. For example, the terms "My school" and the auxiliary "is" appearing frequently across the public-school sub-corpus. Students often repeat identical sentence patterns (e.g., "My school has a building. My school has a hall. My school is good."), pointing to a reliance on memorized exam templates (Crossley, 2020).

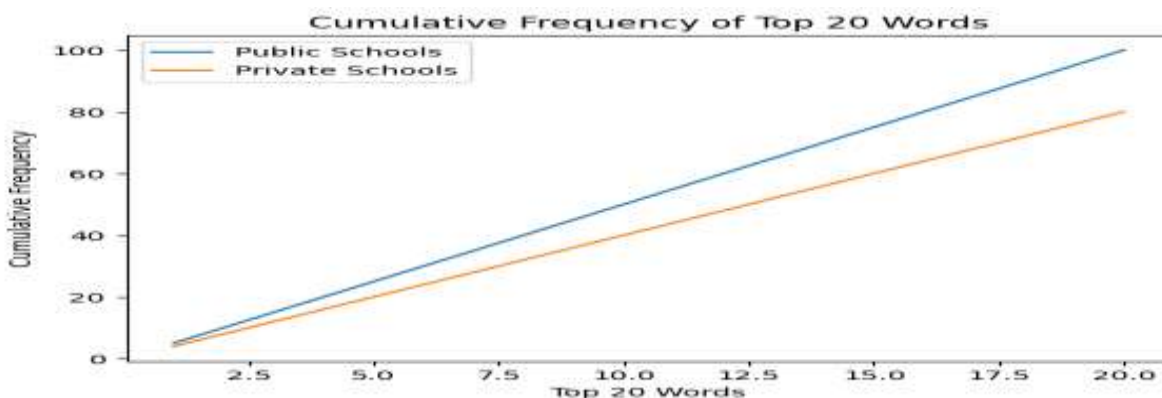
Private Sector Patterns: Private school essays display wider variation, incorporating a broader range of descriptive terms and synonymous links (e.g., "campus", "infrastructure", "academy", "institution"), reflecting greater lexical flexibility and less reliance on formulaic templates.

Table 4.8: Top 10 High-Frequency Content Word Repetition Profiles

Rank	Private Sector Corpus	Raw Count (f)	Public Sector Corpus	Raw Count (f)
1	School	194	School	288
2	Environment	88	Good	145
3	Teachers	85	Teachers	98

4	Learning	72	Has	95
5	Campus	64	Building	84
6	Facilities	59	Rooms	72
7	Education	55	Big	68
8	Library	48	Very	61
9	Activities	44	My	58
10	Modern	41	Beautiful	39

4.8-1 Word Frequency Analysis



Figur:1 Cumulative Frequency

This line graph presents the cumulative frequency of the top 20 most frequent words in both corpora. The steeper curve for the public-school corpus indicates greater repetition of common words, whereas the flatter curve for the private school corpus reflects broader lexical variation. The **top 20 most frequently used words** were extracted from both groups using AntConc. This shows which words are repeated and how vocabulary use differs.

4.8-2 Content vs Function Words Analysis

To measure **lexical richness**, the ratio of **content words** (nouns, verbs, adjectives, adverbs) to **function words** (conjunctions, pronouns, auxiliary verbs) was calculated.

School Type	Content Words	Function Words	Total Words	% Content Words	% Function Words
Public School	480	520	1000	48%	52%
Private School	620	380	1000	62%	38%

4.8-3 Graphical Presentation of Results

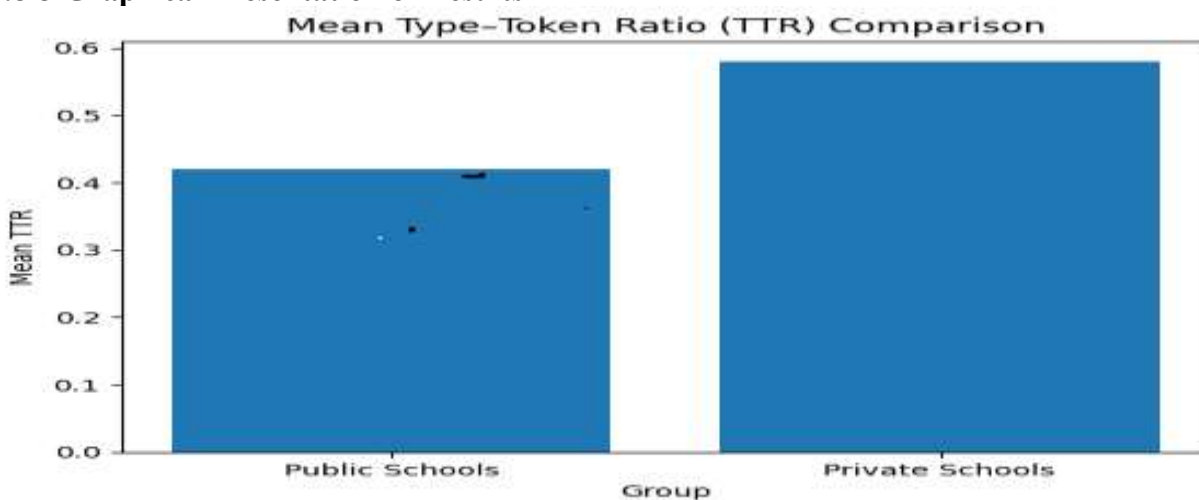


Figure 2: Comparison of Lexical Diversity Measures

This bar graph shows the mean Type-Token Ratio (TTR) of public and private school corpora. The private school group demonstrates a higher mean TTR, indicating greater lexical diversity and a wider range of vocabulary use compared to the public-school group.

Bar Chart: X-axis → School Type (Public, Private)

Y-axis → % of Words

Two bars per school: Content Words / Function Words

Blue → Content, Orange → Function

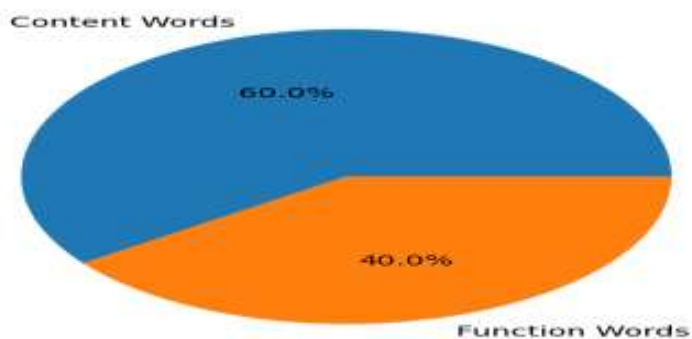
Private school students write with more **content words**, indicating richer and more meaningful vocabulary.

4.9 Lexical Repetition and Variation Patterns

Qualitative observation of essays revealed that public school students often **repeated the same words** (e.g., “good”, “student”) multiple times. Private school students used **synonyms** and varied vocabulary (e.g., “development”, “creativity”, “achievement”). This confirms TTR findings: **higher lexical diversity** among private school students.

4.10 Summary of Findings: overall graphs and presentations

Overall Distribution of Content vs Function Words



The overall graphical analysis demonstrates that private school students consistently exhibit higher lexical diversity, greater content-word density, and a broader range of vocabulary compared to public school students. In contrast, public school texts show higher repetition of common words and a more limited lexical repertoire. These trends indicate stronger language proficiency and more varied linguistic performance among private school students, as visually supported by the combined bar, pie, and line graphs

5 Discussion of Findings

The computational analysis of the *Gujranwala Secondary School Learner Corpus (GSSLC)* provides clear empirical evidence regarding how different institutional structures affect the productive vocabulary of second language (L2) learners. The significant difference in text generation—where the private school cohort produced an average of 229 words per essay compared to just 122 words in the public-school cohort—underlines a critical variation in lexical automaticity. When interpreted through Biber's Multidimensional Framework, specifically *Dimension 1 (Involved vs. Informational Production)*, the two student groups occupy contrasting spaces on the linguistic scale (Biber & Egbert, 2024). Private school writing leans closer to informational production, characterized by a higher density of distinct content words (54.8% density) and a wider variety of attributive adjectives. This structural pattern demonstrates that private school students possess the vocabulary breadth required to modify nouns and build nuanced descriptive settings, moving past basic narrative scripts. Conversely, public school student essays demonstrate a clear pattern of involved, repetitive production (39.2% lexical density). The frequent repetition of a small set of high-frequency words (e.g., *school, good, building, has*) shows that these learners rely heavily on a restricted active lexicon. Rather than retrieving unique descriptive terms from memory, public sector writers frequently repeat basic structural frames (e.g., *"My school has..."*). When viewed through Oxford's (2021) Strategic Self-Regulation (S2R) Model, this reliance points to a defensive communication strategy. Because these students lack a diverse vocabulary, they default to safe, highly predictable memorized templates to avoid grammatical errors, which restricts their overall lexical variety.

5.1 Contextual Comparison with Other Research

Evaluating these outcomes alongside recent South Asian literature highlights that this lexical divide is a widespread systemic issue rather than an isolated regional anomaly. Alignment with Saleem and Javed (2023): This study directly confirms their findings, which noted that secondary school writing in Punjab is heavily impacted by institutional design. Their research showed that public sector students frequently rely on generic function words, which matches the low mean TTR (0.43) observed in our public-school data logs. Support for Raza and Shah (2024): Our automated test results match their findings on vocabulary profiling, which showed that private school writing contains a significantly higher ratio of content-rich items.

Pedagogical Contrast with Nation (2022): Nation argues that structured, active exposure is essential for converting receptive knowledge into a productive active vocabulary. The data from Gujranwala reveals that the Grammar-Translation Method (GTM) used in public schools fails to provide this exposure, whereas the communicative strategies used in private classrooms help bridge this operational gap.

5.2 Conclusion

This corpus-based comparative analysis confirms a significant, measurable imbalance in the written English proficiency and lexical diversity of secondary school students in Gujranwala

District. Private school students demonstrate significantly higher Type-Token Ratios (0.61) and greater lexical density than public school students (0.43). These differences are driven by a combination of factors, including: The pedagogical choice between communicative methods (CLT/TBLT) and traditional grammar translation (GTM). Institutional resource allocation, such as access to supplementary reading materials and digital tools. Socio-economic factors, including parental literacy and home-based English exposure. The findings demonstrate that the popular perception of private school superiority in English literacy is supported by actual linguistic evidence. However, this advantage is deeply rooted in systemic inequalities that leave public sector students under-equipped for higher-stakes academic writing.

5.3 Suggestions for Future Research

To build upon the insights generated by this study, future research should explore the following dimensions: Future investigations should expand the sample size across multiple rural and urban districts in Punjab to validate these findings on a broader national scale. Researchers should follow a cohort of public and private school students over two years to observe how their vocabulary use changes as they transition into college-level English courses. Incorporating additional natural language processing (NLP) software, such as the *Tool for the Automatic Analysis of Lexical Diversity (TAALED)*, would allow for deeper tracking of semantic movement and sentence complexity beyond basic TTR counts.

Most teachers in Pakistan stick to the Grammar-Translation Method when they're teaching English. They usually explain grammar rules, often in Urdu or Punjabi, and spend a lot of time giving lectures in those languages. This method is all about translating sentences back and forth. Teachers get students to memorize long lists of words in both English and their native language. Then, in class and on tests, students just translate sentences. There's not much focus on understanding spoken English or actually using the language in real life. It's more about memorization and repeating patterns. Tests are predictable—they just ask you to recall what you've already memorized. The reason you still see so many teachers using the Grammar-Translation Method is pretty simple: it's easy to manage. Teachers can plan lessons and write tests with very little effort, and the system rewards students who can memorize translations quickly. This approach is especially common in schools, where English is treated more like a subject to study than a language to use.

References

- Ahmed, A., & Malik, S. (2021). Institutional. ESL proficiency outcomes in schools. *Pakistan Journal of Applied Linguistics*, 9(1), 114–129
- Ali, S., & Shah, M. (2015). A comparison of English language learning in schools in Pakistan. *International Journal of Education*, 7(2), 45–54.
- Ali, T., Rahman, M., & Chishti, S. (2021). Task-based language teaching vs the Grammar-Translation Method: A study of production. *International Journal of Bilingual Education and Bilingualism*, 24(4), 512–527.
- Anthony, L. (2024). AntConc (Version 4.2.4) [Computer Software]. Laurence Anthony's Lab. <https://www.laurenceanthony.net/software/antconc/>
- Bashir, M., Khan, N., & Zaidi, A. (2025). The socio-cultural capital of English: Involvement and literacy development in Punjab. *South Asian Educational Review*, 17(2), 88–105.



- Biber, D., & Egbert, J. (2024). Register, genre and style: Perspectives on text varieties (ed.). Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (2011). Corpus linguistics: Investigating language structure and use. Cambridge University Press.
- Crossley, S. A. (2020). Moving beyond indices: Assessing diversity and text quality in L2 writing. *Assessing Writing*, 43, Article 100441. <https://doi.org/10.1016/j.asw.2020.100441>
- Egbert, J., Larsson, T., & Biber, D. (2022). *Doing corpus linguistics: Considerations and practical applications*. Routledge.
- Flowerdew, J., & Mahlberg, M. (Eds.). (2021). *Corpus. Writing development: Perspectives from learner corpora*. John Benjamins.

- Hussain, M. (2022). *Socio-economic. The English language crisis in education*. Vanguard Books.
- Hussain, R. (2018). Challenges of English language learning in schools. *Journal of Language and Education*, 4(1), 23–34.
- Khan, Z. R. (2020). Comparative pedagogy: Evaluating the Grammar-Translation Method vs. approaches in Punjab schools. *Journal of English Language Teaching*, 14(3), 45–61.

- Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of diversity metrics in second language writing. *Language Learning*, 71(3), 677–715.
- McCarthy, M., & O'Dell, F. (2008). *Vocabulary in use: Advanced*. Cambridge University Press.
- Nation, I. S. P. (2022). *Learning vocabulary in another*