

**What Errors Say: Errors and influencing factors on written interlanguage in
argumentative essays of Colombian EFL pre-service teachers**

Juan Carlos García Granada

(202022257)

Sebastián Paz Granobles

(202028319)

Director

Orlando Chaves Varón

UNIVERSIDAD DEL VALLE
ESCUELA DE CIENCIAS DEL LENGUAJE
DEPARTAMENTO DE LENGUAS Y CULTURAS EXTRANJERAS
LICENCIATURA EN LENGUAS EXTRANJERAS
2025

TABLE OF CONTENTS

ABSTRACT	7
INTRODUCTION	10
1.1. Preamble	10
1.2. Significance of the study	12
LITERATURE REVIEW	14
2.1. Conceptual framework	14
2.1.1. Errors	14
2.1.2. Interlanguage	15
2.1.3. Corpus Linguistics	16
2.1.4. Learner corpora	18
2.1.5. Computer-based Error Analysis	19
2.1.6. Socioeconomic Conditions and EFL Writing	20
2.1.7. Motivation and Enjoyment in EFL Writing	23
2.2. Previous studies	25
2.2.1. International research	25
2.2.1.1. Studies on error analysis	25
2.2.1.2. Studies on socioeconomic conditions	30
2.2.1.3. Studies on writing enjoyment	33
2.2.2. National research	37
2.2.3. Local research	40
RESEARCH METHODOLOGY	45
3.1. Research approach and design	45
3.2. Research site and participants	46
3.2.1. Research site	46
3.2.2. Participants	47
3.3. Data collection methods, procedures and analysis	48
3.3.1. Data collection methods	48
3.3.1.1. Students' written compositions	48
3.3.1.2. Survey	50
3.3.2. Procedures and analysis	54
3.3.2.1. Error analysis	54
3.3.2.2. Text alignment	58
3.3.2.3. Peer review	60
3.3.2.4. Correlations	62
3.4. Validity and reliability	65

3.5. Ethical considerations	66
RESULTS AND DISCUSSION	67
4.1. Predominant errors in the written compositions	67
4.1.1. Results	67
4.1.1.1. Analysis by general categories	67
4.1.1.2. Analysis by subcategories	71
4.1.2. Interpretation and discussion of the results	76
4.2. Correlation between the participants' social strata and the frequency and types of interlanguage errors	79
4.2.1. Results	79
4.2.2. Interpretation and discussion of the results	83
4.3. Correlation between participants' level of writing enjoyment and the frequency and types of errors	86
4.3.1. Results	86
4.3.2. Interpretation and discussion of the results	89
CONCLUSION	91
REFERENCES	98
APPENDIX	105

List of Figures

Figure 1. Example of a Trial Test	50
Figure 2. Title and Introduction of the Survey	51
Figure 3. Question on Liked Aspects of the English Language	52
Figure 4. Question on the Socioeconomic Stratum of the Participants	53
Figure 5. Example of Error Tagging in a Plain Text Format	58
Figure 6. Example of Text Alignment Process in Concord Tool	59
Figure 7. Categorized Data on Texts and Found Errors in Google Sheets	60
Figure 8. Alignment of Total Errors for Professor's Review	61
Figure 9. Data for Correlations in Google Sheets	63
Figure 10. Data Imported into JASP	64
Figure 11. Descriptive Statistics and Correlations Carried Out in JASP	65
Figure 12. Errors per Category in the Corpus	68
Figure 13. Number of Grammar Errors in the Corpus	69
Figure 14. Percentages of Form Errors in the Corpus	70
Figure 15. Percentages of Punctuation Errors in the Corpus	71
Figure 16. Errors per Subcategory in the Corpus	72
Figure 17. Students' Socioeconomic Strata	80
Figure 18. Students' Writing Enjoyment	87

List of Tables

Table 1. Error Categories and Their Abbreviations	55
Table 2. Examples of Predominant Error Types and Their Corrections	75
Table 3. Descriptive Statistics for Students' Social Strata	81
Table 4. Shapiro-Wilk Test of Normality for Socioeconomic Strata and Total Errors	82
Table 5. Pearson's Correlation Coefficient Between Errors and Socioeconomic Strata	83
Table 6. Descriptive Statistics for Students' Writing Enjoyment.	87
Table 7. Shapiro–Wilk Test of Normality for Writing Enjoyment and Total Errors	88
Table 8. Spearman's Rank Correlation Coefficient Between Errors and Writing Enjoyment	88

Authors' Note

During the writing of this thesis, the authors used the AI-powered writing assistants Grammarly and LanguageTool for correction and revision of grammar, spelling, and punctuation, as well as to verify the clarity and concision of the writing.

ABSTRACT

Multiple studies have demonstrated that error analysis constitutes a pivotal instrument in identifying the challenges and areas of difficulty learners face during the acquisition process of a new language. This process has also been proven to be immensely valuable for teaching. However, it has not been widely studied in university contexts in Colombia. The primary objective of the present study is to characterize the written interlanguage in argumentative essays composed by a group of 49 eighth-semester English as a Foreign Language (EFL) pre-service teachers (correlations established with a 30-participant subgroup) at the Universidad del Valle in Cali, Colombia. Furthermore, this study aims to contribute to the ongoing Colombian national corpus project, which investigates the same linguistic phenomenon across various universities within the country.

To achieve these objectives, this study employed a quantitative approach, using the Error Tagging Manual 1.2 developed by the University of Louvain for systematic error categorization, and an online survey to gather information on sociocultural factors that can influence these errors. The statistical analysis of the data was conducted using the software JASP 0.18.1. The findings indicate that the primary errors observed in the written productions of pre-service EFL teachers can be classified into five main types: Form (FS), Missing Punctuation (QM), Grammar Article (GA), Lexicon (LS), and Infelicities (inappropriate or awkward expressions, Z). Against what was expected, it was found that correlations between errors and variables such as socioeconomic status and writing enjoyment are not statistically significant. The study contributes empirical information to inform pedagogical strategies to improve the written interlanguage of intermediate EFL learners.

Keywords: argumentative essays, Computer-based Error Analysis, written interlanguage, intermediate EFL learners, pre-service teachers, sociocultural factors in EFL learning.

RESUMEN

Múltiples estudios han demostrado que el análisis de errores constituye una herramienta fundamental para identificar los desafíos y las áreas de dificultad que enfrentan los aprendices durante el proceso de adquisición de una nueva lengua. Este proceso también ha probado ser inmensamente valioso para la enseñanza. Sin embargo, no ha sido ampliamente estudiado en contextos universitarios colombianos. El objetivo principal del presente estudio es caracterizar la interlengua escrita en ensayos argumentativos elaborados por un grupo de 49 profesores de inglés como lengua extranjera (ILE) en formación (correlaciones establecidas con un subgrupo de 30 participantes), de octavo semestre en la Universidad del Valle en Cali, Colombia. Adicionalmente, este estudio busca contribuir al proyecto nacional colombiano de corpus en curso, que investiga el mismo fenómeno lingüístico en diversas universidades del país.

Para alcanzar estos objetivos, se empleó un enfoque cuantitativo, utilizando el Error Tagging Manual 1.2 desarrollado por la Universidad de Lovaina para la categorización sistemática de errores, y una encuesta en línea para recopilar información sobre factores socioculturales que pueden influir en dichos errores. El análisis estadístico de los datos se realizó mediante el software JASP 0.18.1. Los hallazgos de este estudio indican que los principales errores observados en las producciones escritas de los profesores de ILE en formación pueden clasificarse en cinco tipos principales: Forma (FS), Puntuación Faltante (QM), Artículo Gramatical (GA), Léxico (LS) e Infelicidades (vocabulario o expresiones inapropiadas, Z). Contra lo esperado, se encontró que las correlaciones entre errores y variables como el estatus socioeconómico y el gusto por la escritura no son estadísticamente significativas. El estudio contribuye datos empíricos que pueden servir de base para el diseño de estrategias pedagógicas para la enseñanza de la escritura en inglés como lengua extranjera.

Palabras clave: ensayos argumentativos, análisis computarizado de errores, interlengua escrita, estudiantes intermedios de ILE, profesores en formación, factores socioculturales en el aprendizaje de ILE.

RÉSUMÉ

De nombreuses études ont démontré que l'analyse d'erreurs constitue un instrument clé pour identifier les défis et difficultés que rencontrent les apprenants durant le processus d'acquisition d'une nouvelle langue. Ce processus s'est également révélé extrêmement utile pour l'enseignement. Toutefois, il n'a pas été largement étudié dans les contextes universitaires au niveau local. L'objectif principal de cette étude est de caractériser l'interlangue écrite dans des essais argumentatifs produits par un groupe de 49 professeurs d'anglais langue étrangère (ALE) en formation en huitième semestre (des corrélations établies avec un sous-groupe de 30 participants) à l'Universidad del Valle de Cali, en Colombie. De plus, cette étude vise à contribuer au projet national colombien de corpus en cours, qui examine le même phénomène linguistique dans diverses universités du pays.

Pour atteindre ces objectifs, cette étude a employé une approche quantitative, utilisant l'Error Tagging Manual 1.2 développé par l'Université de Louvain pour la catégorisation systématique des erreurs, ainsi qu'une enquête en ligne pour recueillir des informations sur les facteurs socioculturels pouvant influencer ces erreurs. L'analyse statistique des données a été réalisée avec le logiciel JASP 0.18.1. Les résultats de cette étude indiquent que les principales erreurs observées dans les productions écrites des professeurs en formation d'ALE peuvent être classées en trois catégories majeures : Forme (FS), Ponctuation Manquante (QM), Article Grammatical (G), Lexique (LS) et Imperfections (Z). Contrairement aux attentes, on a constaté que les corrélations entre erreurs et variables, telles que le statut socio-économique et le plaisir d'écrire, ne sont pas statistiquement significatives. Cette étude apporte des informations empiriques permettant d'éclairer les stratégies pédagogiques visant à améliorer l'interlangue écrite des apprenants intermédiaires d'anglais langue étrangère.

Mots-clés : essais argumentatifs, analyse d'erreurs assistée par ordinateur, interlangue écrite, apprenants intermédiaires d'ALE, professeurs en formation, facteurs socioculturels dans l'apprentissage d'ALE.

INTRODUCTION

1.1. Preamble

The concepts of interlanguage and error analysis are intrinsically linked and serve as valuable tools for comprehending the process of second language acquisition (Corder, 1981; Selinker, 1972). In this approach, errors are not viewed as obstacles, but rather as essential steps in the language-learning journey (Corder, 1981). By examining these errors, educators can gain profound insights into a learner's current level of language knowledge and the strategies they employ during the learning process. Corder's perspective suggests that when a learner makes errors, it indicates that they possess a personal 'competence' or a systematic knowledge base, as Chomsky would put it (Corder, 1981). In other words, learners have a well-defined, albeit evolving, personal grammar that forms the foundation of their linguistic expressions. It is now well established from a variety of studies that error analysis can be an important tool to determine the currency of errors and determine if they are caused or not by transfer from the learners' first language (Hoyos, 2016). It should also be noted that most studies have been descriptive, focused on identifying, classifying, and describing students' errors in their linguistic productions (Lemos, 2022; Pardo, 2021).

Despite the relevance of error analysis, little is currently known about the relationship between the interlanguage error production and learners' demographic and social variables (Pardo & Quiroz, 2021), especially within the Colombian context of English as a Foreign Language (EFL) pre-service teachers (Chaves, 2011). Pardo and Quiroz (2021) found that sociodemographic factors such as gender and social strata were not statistically significant in explaining error patterns. Their findings support theories suggesting that women tend to perform

better in language learning, while contradicting theories by revealing that learners from lower social strata make fewer mistakes than learners from higher social strata. These mixed results highlight the scarcity of local studies on this topic and continue the need for further exploration of how individual and sociocultural variables interact with linguistic performance in Colombian EFL contexts.

To address this gap, the present study seeks to examine the most common errors and two sociocultural factors in the written interlanguage of 8th-semester pre-service teachers of the Foreign Languages program at the Universidad del Valle, in Cali, Colombia. This is to be reached through both an error analysis of students' argumentative written productions and a correlational analysis of students' social strata and writing enjoyment. In response to this issue, the present study formulates the following research questions.

What are the most common errors in the written interlanguage of eighth-semester English Foreign Language pre-service teachers at the Universidad del Valle, and is there a correlation between these errors and sociocultural factors?

To guide the study, the following specific questions were formulated.

- What are the predominant errors in the written compositions of eight-semester EFL pre-service teachers from the Universidad del Valle?
- To what extent is there a correlation between the participants' social strata and the frequency and types of interlanguage errors observed in their argumentative essays?
- To what extent is there a correlation between participants' level of writing enjoyment and the frequency and types of interlanguage errors present in their argumentative essays?

1.2. Significance of the study

The present investigation about the written interlanguage of EFL pre-service teachers and its potential correlation with sociocultural variables has significant importance for the local academic context of Colombian teacher training, the pedagogical practices within EFL education, and the general theoretical models of Second Language Acquisition (SLA).

First, this study is significant for addressing a major gap in the local Colombian EFL context, specifically related to advanced pre-service teachers. By analyzing their errors, it is useful as a key diagnostic tool to improve teacher training, writing instruction, and curriculum development at Universidad del Valle, given that pre-service EFL teachers are not only language learners but also future important agents in the educational system. Due to the latter, understanding specific challenges and characteristics of their interlanguage is essential. Furthermore, it contributes to a broader national corpus project on the same phenomenon, allowing for collaborative and comparative analysis of English learning across Colombian universities.

Second, the theoretical significance of this research for the SLA field lies mainly in its aim of going beyond the mere error description. While most of the existing research on error analysis is predominantly descriptive, focusing on the identification and classification of errors, this study seeks to examine the correlations between linguistic errors and sociocultural variables. Moreover, it tests the validity of theories like Bourdieu's cultural capital within a specific EFL context by examining the influence of socioeconomic status on writing performance. Additionally, this study revises motivational models in the ELF context by exploring the link between writing enjoyment and accuracy, challenging assumptions such as considering that

enjoyment directly translates into better performance. These perspectives enrich the understanding of interlanguage as a system influenced not only by cognitive factors but also by the complex interaction of sociocultural and individual variables.

Finally, this study also offers some practical and pedagogical contributions in the field of EFL teacher education by identifying the most frequent errors in advanced learners' written compositions and serving as a source of information on instructional priorities in the local context. For instance, educators from the academic program could use the insights resulting from this study to develop targeted lessons or didactic materials to address specific needs in learners. Moreover, with the exploration of the role of writing enjoyment, this study contributes to the understanding of the affective dimensions of learning to write in a foreign language, which could also be useful for creating a classroom environment that not only addresses linguistic aspects of learning but also fosters a positive writing motivation. Now, even if sociocultural variables were not influential in linguistic performance when writing in this specific context, this study would recommend the development of more sensitive pedagogical approaches that recognize the importance of the diversity of backgrounds and motivations that learners bring to their learning process, which would ultimately contribute to improvements in equity and efficacy in language teaching.

In summary, this research is significant because it provides needed empirical data on a local context, deepening existing theoretical models, and offering foundations to enhance pedagogical practices within a degree of English language teachers in Cali, Colombia.

LITERATURE REVIEW

Given that the objective of this study is to find out the most common errors in the written interlanguage of a group of students and examine if there is a relation with two sociocultural factors, the following section will present an overview of, firstly a conceptual framework that will serve to understand the main notions and theories regarding the subject of study, and secondly the empirical research conducted at the local, national and international level.

2.1. Conceptual framework

For this study, the constructs that are necessary to provide an understanding of the research problem are the following: errors, interlanguage, corpus linguistics, learner corpora, and computer-based error analysis. These concepts are integrated within a sociocultural view of writing as a socially and culturally situated phenomenon. This theoretical perspective drives our aim to examine the correlation between a linguistic phenomenon (EFL written errors) and individual and social factors (writing enjoyment and socioeconomic status).

2.1.1. Errors

Errors have traditionally had negative connotations, seen as something to be avoided at all costs. This negative view of errors comes from old language theories, such as language-centered views or structuralism, in which the teaching process focused on the language system itself (grammar rules, vocabulary and sentence structure) and learning focused on imitating correct forms and correcting errors. Corder (1981) stated a different approach, differentiating between mistakes and errors. For this author, the errors of performance are slips that can happen to anyone (even in their mother tongue), and for him, these are mistakes. On the

other hand, for Corder, the term error refers to the systematic errors committed by a learner, which allows us to reconstruct their knowledge of the language they are learning.

For Corder (1981), errors are significant in three different ways. For the teacher, errors are a good way to see how much towards the goal the learner has progressed and what is missing for them to learn. For the researcher, they provide evidence of how language is acquired and the strategies or procedures learners are employing. And for the learner himself, the errors serve as a device they use in order to learn, since this is how they can test their hypothesis about the language they are learning. Corder also states that people learning their first language have an unlimited number of hypotheses about the nature of the language they are learning, while a person learning a second language has only one hypothesis; they can only compare that new language with the one they already know and wonder if the systems of that new language are the same as the ones they already understand. The evidence to support this is the large number of errors related to the mother tongue (cross-linguistic transfer), since the learner is investigating the systems of the new language, and this is why errors should be seen as evidence of learners' learning strategies. Taking into account this focus on the importance of errors and the interference of hypotheses related to learners' mother tongue, the concept of *interlanguage* arises, as reviewed in the following subsection.

2.1.2. Interlanguage

To introduce readers to the concept of *interlanguage*, it is necessary to explore the contributions of scholars who paved the way for its development. Before the introduction of the *interlanguage* concept, Nemser (1971) introduced the idea of *Approximative Systems*. Nemser's theory suggests that language learners construct unique linguistic systems that neither precisely

mirror the target language nor their native language. Instead, these systems represent approximations that evolve as learners progress in their language acquisition journey.

Years later, the term *interlanguage* was first introduced by Selinker in 1972, describing it as a system based on the observable output of a second language (L2) learner. This theory of second language learning, proposed by Selinker, was based on a *latent psychological structure* (Lenneberg, 1967), a pre-existing cognitive structure that is activated when a learner tries to learn a second language. These attempts to use the target language (TL) are crucial in determining learners' success. To develop his theory, Selinker compared the utterances of learners and native speakers of the TL, revealing the existence of a system based on the observable output.

Several authors have provided definitions and insights into interlanguage, including Corder (1981), who highlighted the significance of errors in shaping the development of a new linguistic system during L2 acquisition. Corder also introduced the concept of *transitional competence*, referring to the evolving body of knowledge possessed by learners about the target language. It is the responsibility of applied linguists to investigate and understand this phenomenon. In recent years, several researchers have found that a course of action to study interlanguage is through the use of corpus linguistics, a methodology that can help in the analysis of large collections of data.

2.1.3. Corpus Linguistics

Corpus linguistics has emerged as a vibrant and influential field within the broader domain of linguistics, offering a data-driven approach to the study of language. It is characterized by the systematic analysis of large collections of authentic texts, known as corpora,

to uncover patterns of language use and variation. This empirical approach has revolutionized the understanding of language, providing insights into grammar, lexis, discourse, and language change. According to Kennedy (1998), corpus linguistics can be defined as the study of language based on the analysis of large electronic collections of text, or corpora, which are carefully compiled and tagged with linguistic information, allowing for the systematic investigation of language patterns and their usage across different genres, registers, and time.

The roots of corpus linguistics can be traced back to the mid-20th century, with the development of electronic text corpora and the emergence of computational linguistics. However, it was not until the 1960s and 1970s that corpus linguistics became more established as a distinct field, driven by scholars such as John Sinclair, W. Nelson Francis, and Henry Kučera (Villayandre, 2008). Corpus linguistics has made significant contributions to the field of language studies, providing valuable insights into the processes of language acquisition, language variation, and language pedagogy (Yakut, 2022). By analyzing large collections of learner language data, corpus linguists have shed light on the developmental trajectories of second language learners, identifying patterns of error, grammatical development, and vocabulary acquisition (McEnery & Hardie, 2011; Yakut, 2022).

As claimed by Lozano and Mendikoetxea (2013), corpus-based research has also informed the development of more effective teaching materials and methodologies. By examining the authentic use of language in different contexts, teachers can identify the most common and useful vocabulary, grammatical structures, and discourse patterns of learners at particular stages of their second language acquisition process and the developmental limitations that constrain the production in a second language (L2). This information can then be used to

design more targeted and effective language instruction adapted for second language learners from a specific context.

Now, to further explore the concepts related to this study and the observed insights about corpus linguistics, it is necessary to define the concept of *learner corpora*, given that it focuses specifically on the linguistic output of language learners and delves into the dynamic process of language acquisition and development.

2.1.4. Learner corpora

Learner corpora, emerging in the late 1980s and early 1990s, represent a dynamic field within corpus linguistics. According to Granger (2012) and Bennett (2010), learner corpora are a kind of specialized corpora that consist of electronic compilations of texts (written texts and/or oral transcripts) produced by individuals learning a language, generally focusing on second or foreign language learners (Granger, 2012; Bennett, 2010). As stated by Granger (2012), this type of corpus serves two main purposes: contributing to Second Language Acquisition theory by offering insights into aspects such as interlanguage, and contributing to the development of pedagogical tools that better address the needs of language learners (Granger, 2012).

Additionally, it is worth noting that the term ‘learner corpus’ encompasses data from foreign language learners, excluding the population whose target language is their first language or an institutionalized additional language. A widely known example of a learner corpus is the International Corpus of Learner English, which contains written compositions authored by English language learners with 14 different native languages (Granger, 2003).

2.1.5. Computer-based Error Analysis

The field of learner corpus analysis has evolved with time; what was previously made by hand has been automated along with technological advances. Granger (2008) highlights two prominent approaches for analyzing learner data: constructive interlanguage analysis (CIA) and computer-aided error analysis (CEA). CIA compares learner data with native speaker data (L2 vs. L1) or different types of learner data (L2 vs. L2). While most CIA studies rely on unannotated learner corpora, a few have employed part-of-speech tagged corpora to examine the frequency of grammatical categories or sequences. On the other hand, CEA involves examining learner errors using learner corpora. In these corpora, error tags and possible corrections are added using a specialized editing tool. Unlike traditional error analysis, CEA does not isolate errors from the original texts but studies them within the context of correct use and over- or underuse. This paradigm shift reflects the evolution of learner corpus analysis, adapting to technological progress for more comprehensive insights into language learning patterns.

To complement the aforementioned, a recent literature review written by Mariappan et al. (2022) provides valuable insights into the strengths and limitations identified in various research studies involving Computer-Based Error Analysis (CBEA). Among the strengths, CBEA has been noted for its ability to provide solutions, ensure accuracy and precision in error detection, offer instant analysis, and significantly reduce the workload for teachers. These attributes contribute to the efficiency and effectiveness of CBEA in aiding ESL and EFL learners. However, it is crucial to acknowledge certain limitations associated with CBEA, such as its challenges in analyzing higher-level errors (for instance, problems in recognizing errors in long and complex sentences, misinterpretation of idioms and collocations, and avoidance of semantic issues), the need for multiple software packages to cover diverse error types, and its limitation in

identifying content and coherence errors. These considerations highlight the nuanced landscape of CBEA's strengths and limitations in language learning contexts.

For the present research, the term CEA will be used as equivalent to Computer-Based Error Analysis (CBEA) for practical purposes, since there is no major difference between the two concepts.

2.1.6. Socioeconomic Conditions and EFL Writing

Many theorists in sociolinguistics and education have theorized about the connection between a learner's socioeconomic background and their writing proficiency in English as a Foreign Language (EFL). Regarding this matter, Bernstein's (2003) sociolinguistic code theory offers an elemental explanation by linking linguistic behavior to social class. He proposed two linguistic codes: restricted and elaborated. These codes are associated respectively with working-class and middle-class speakers. Restricted codes are considered context-dependent and more implicit, often limiting learners' linguistic flexibility, lexical variety, and syntactic complexity, all essential elements for academic writing. Consequently, EFL learners from lower SES backgrounds, more familiar with restricted codes, are likelier to find difficulties when required to produce elaborated, decontextualized, and explicit written compositions, which are typical of academic settings. Thus, SES indirectly shapes EFL writing competence by influencing students' linguistic resources and their ability to adapt to formal and academic writing norms (Bernstein, 2003).

While Bernstein viewed linguistic differences through a sociological lens that implied a deficit in working-class language use, Labov (1990) opposed this interpretation through his Theory of Linguistic Variation and Social Stratification. Labov demonstrated that all social

dialects, including those associated with low-SES groups, follow consistent grammatical and phonological patterns rather than representing linguistic deficiencies. His perspective reformulates the subject of writing performance, considering that challenges of lower SES learners in EFL writing emerge not from linguistic inferiority but from mismatched sociolinguistic expectations between home and academic discourse. Thus, their nonstandard linguistic variations may transfer into EFL writing, producing deviations that are perceived as errors in formal contexts (Labov, 1990). Labov's perspective, therefore, emphasizes that the educational system's preference towards certain language forms privileges middle-class sociolects while undervaluing and marginalizing working-class language varieties.

Similarly, the theory of capital proposed by Bourdieu (1986) can provide a more comprehensive explanation of the impact that social inequality has on EFL writing by examining the unequal distribution of cultural capital, including aspects like knowledge, language practices, and educational dispositions that are mainly acquired through family socialization. For instance, learners from higher SES backgrounds are more likely to possess a more significant exposure to literacy practices and academic language, enabling them to internalize linguistic and rhetorical conventions valued in formal writing. Conversely, EFL learners from a lower SES have limited cultural and economic capital, which means that they have to face more difficulties to access the linguistic resources and learning environments that lead to developing proficient skills in EFL writing. In brief, Bourdieu's framework rationalizes the fact that linguistic proficiency and writing performance are not merely cognitive achievements but social products shaped by access to symbolic and material resources.

Complementing these sociological views, Gardner's (2006) socio-educational model of second language acquisition (SLA) and Ellis's (1994) discussion of social factors in SLA highlight that the students' socioeconomic background influences EFL writing indirectly. This influence can be evidenced through motivation, attitudes, and learning opportunities. On the one hand, Gardner argued that learners from more privileged environments often demonstrate stronger instrumental and integrative motivation, which sustains persistence in complex tasks such as writing in a foreign language. Similarly, Ellis (1994) emphasized that SES determines learners' access to quality input, instruction, and opportunities for interaction, all of which shape language-related dimensions of writing, such as accuracy, syntactic complexity, and lexical range. Thus, the effects of SES extend beyond linguistic knowledge, influencing learners' affective engagement, exposure, and ultimately, their written performance.

In synthesis, the abovementioned theoretical perspectives align with the idea that EFL writing performance is influenced and shaped by socioeconomic background through intertwined linguistic, cultural, and motivational mechanisms. While Bernstein and Bourdieu emphasize structural and cultural inequalities in language access and capital, Labov, Gardner, and Ellis provide complementary views that reinterpret linguistic variation as socially patterned and highlight the facilitating roles of motivation and exposure. Together, these theories suggest that differences in EFL writing outcomes are not the result of individual deficits but reflect a broader issue in social stratification processes that govern linguistic opportunities and access to education.

2.1.7. Motivation and Enjoyment in EFL Writing

Theories of motivation and affect offer a comprehensive framework for understanding the psychological processes that influence writing performance in English as a Foreign Language (EFL). Boscolo and Hidi (2007) conceptualize writing motivation as a complex multidimensional construct encompassing interest, self-efficacy, and self-regulation, which are all affected by social and contextual factors. They argue that students' motivation to write is shaped by both intrinsic and extrinsic reasons (for instance, the desire to improve, personal interests, or meeting others' expectations), and that these reasons directly influence aspects as the effort, engagement and persistence that students show in writing tasks. The emotional component of writing, particularly enjoyment, is a crucial factor in maintaining writer motivation as it encourages deeper cognitive involvement and more effort, while negative emotions or affective states such as anxiety or boredom can reduce learners' performance (Boscolo & Hidi, 2007).

In this way, affect and motivation interact in a continuous, dynamic cycle that influences writing: On the one hand, positive emotions and self-confidence promote greater engagement, which leads to careful revisions, attention to language accuracy, and creative experimentation that improve their skills. On the other hand, negative emotions like anxiety or boredom can limit mental energy, making it harder to concentrate. This often results in rushed or shallow writing and a higher chance of mistakes.

From a motivational perspective, the social-cognitive theory proposed by Bandura (1986, 1997) has been particularly important in explaining individual differences in writing outcomes. Within this theory, self-efficacy is the belief in one's own ability to perform a specific task, and it is seen as a key motivational factor for writing performance. Studies carried out by Pajares,

Valiante and Cheong (2006) demonstrate that students with higher writing self-efficacy usually demonstrate a stronger self-regulation, greater persistence, and more positive motivational patterns and attitudes. Also, they report less anxiety when writing and significantly higher enjoyment when composing texts. In EFL contexts, self-efficacy interacts with linguistic competence. For instance, learners who perceive themselves as good writers are more likely to view errors as part of the learning process instead of as failures, facilitating continuous improvement in language use. For this reason, the quality and accuracy of written productions tend to be a result influenced by different aspects such as enjoyment and motivation, as well as self-efficacy.

Theories of writing motivation emphasize that enjoyment of writing plays a very important role in sustaining engagement and improving performance. Writing is not only a cognitive task but also a socially mediated activity that acquires meaning through interaction, feedback, and authentic communication (Boscolo & Hidi, 2007). Learning environments that encourage collaboration, autonomy, and personally meaningful topics are helpful factors to enhance students' intrinsic motivation and enjoyment of writing. As shown by De Bernardi and Antolini (2006), these conditions increase learners' involvement and satisfaction, leading to a greater attention to form and content and, consequently, resulting in the commission of fewer linguistic errors. Moreover, affective and social-cognitive theories highlight that enjoyment can function as both a cause and an outcome of writing success. On the one hand, positive emotional experiences may reinforce self-efficacy and persistence, and on the other hand, successful writing performances can generate further enjoyment and confidence. In EFL contexts, where language limitations are commonly a cause of anxiety and limited fluency, promoting writing enjoyment can mitigate negative emotions and their related consequences. Therefore, enjoyment

is not a secondary factor but a key motivational element that reinforces effort, self-regulation, and accuracy in EFL writing tasks, ultimately contributing to the overall performance.

To finish this theoretical framework, all the constructs previously mentioned serve as a foundation for understanding the studies that are about to be analyzed as empirical research in the following section.

2.2. Previous studies

This literature review examines interlanguage and error analysis studies in diverse contexts. Through a thorough examination of both international and national studies, the review critically assesses the methods, participants, designs, instruments, findings, and recommendations put forth by researchers in this field. Furthermore, it endeavors to enhance comprehension of the implications of these concepts, with the ultimate goal of addressing knowledge gaps in English writing for EFL preservice teachers.

2.2.1. International research

2.2.1.1. Studies on error analysis

Error analysis has been widely studied across the world, analyzing the students' performance with different typologies and in different learning stages. The selected studies have investigated various aspects of errors committed on written compositions, employing diverse methodologies to identify, categorize, and analyze errors made by students.

Díez-Bedmar (2011) conducted an empirical analysis of Spanish pre-university students' written compositions for the University Entrance Examination (UEE). The study aimed to

achieve two primary objectives: first, to provide an updated overview of the main challenges that students face when using English to write compositions for the UEE, and second, to employ a methodology that ensures comparable results, addressing limitations in previous research. The learner corpus used for the analysis consisted of 1,406 compositions at the University of Jaén, focusing on the topic ‘Where, outside Spain, would you go on a short pleasure trip?’. This research was qualitative, and to analyze the information found in the writing compositions, the researcher employed the *UCL Error Editor* (Hutchinson, 1996) and the *Error Tagging Manual, version 1.1*. (Dagneaux, Denness, Granger & Meunier, 1996) designed by the Centre for English Corpus Linguistics at the Université Catholique de Louvain. Of the identified errors, totaling 5,811 in 34,403 words, the most prominent errors were related to grammar, followed by lexical and formal issues. The study’s limitations include the specificity of the sample and topic, suggesting the need for replication with diverse samples and topics. Despite the specificities of this research, the study contributes valuable insights into language proficiency issues among Spanish pre-university students, highlighting the need for the design of courses and materials to help the students improve their written compositions. This research also emphasizes the importance of consistent methodologies for future research in this domain.

Similarly, in a descriptive study, Garrido and Rosado (2012) examined the errors made by forty-eight Spanish speakers studying to become EFL teachers at Universidad San Sebastián, in Concepción, Chile. The main objectives of the research were to identify errors related to the use of tenses and aspects in English and establish a hierarchy of difficulty. This quantitative and descriptive study applied Corder’s error analysis model, comparing learner utterances with those of native speakers. Also, error categories were defined based on Dulay, Burt, and Krashen’s (1982) surface taxonomy. Overt errors, which are easily noticeable and can be explicitly

identified, were observed. These included the simple present, past progressive, simple future, and present perfect, with frequencies of 20.37%, 18.5%, 18.5%, and 16.7%, respectively. Notably, covert errors, which are subtle mistakes that may not be immediately apparent to the learner, were identified. The highest frequencies included the present perfect progressive, past perfect, present progressive, and simple future (28.73%, 21.83%, 20.68%, and 14.94%, respectively). These findings offer useful insights for the development of remedial programs aimed at enhancing students' proficiency in utilizing English tenses and aspects. The study underscores the significance of error analysis as a practical tool for comprehending learners' language. By delving into both covert and overt errors, this research introduces a novel approach to error analysis, making a substantial contribution to the field.

Another study that depicts similar results is the one carried out by Alawi (2014), who undertook a descriptive investigation to depict the interlanguage of English language learners and assess their writing competence. The study centered on 105 first-year students at Surigao del Sur State University, in the Philippines, aiming to describe learners' interlanguage by examining errors in their written compositions. This research employed performance assessment, interlanguage analysis, and error analysis as key methodologies. The study focused on written compositions, and after selecting papers, errors were identified, labeled, and measured in terms of error density and frequency. The findings revealed that approximately eight out of ten clauses or sentences contained errors, with verb usage proving to be the most problematic area, constituting 20.5% of the total 6,386 errors. Punctuation and capitalization followed closely at 16.4% and 10.9%, respectively, while issues with articles as modifiers were also identified as problematic. The study emphasizes the potential impact of error frequency on language education, curriculum design, instructional materials, and teaching strategies. Despite being a

descriptive study, this research provides valuable insights into error analysis, showing the calculation of the error density index and offering practical applications for language educators and planners. While the study does not delve into the causes of errors, it serves as a great resource for understanding and addressing language learners' challenges in academic writing.

Another recent study worth highlighting is the one conducted by Alsahafi (2017), who analyzed the types and causes of the spelling, lexical and grammatical errors made by 70 Saudi EFL Foundation Year students in their written compositions. The main objectives were to classify the most frequent errors, their possible linguistic explanations and examine whether increasing learners' attention to language would improve their accuracy, since the students were also faced with a discrete-point test. The study applied two error classification frameworks: a linguistic category taxonomy and a surface strategy taxonomy. Through quantitative and qualitative analyses, the researcher found that students demonstrated significantly better performance in orthographic accuracy (particularly in spelling vowels and silent letters) when their attention was directed to form. However, improvements were not observed at the lexical or grammatical levels. The findings suggest that many orthographic errors come from the transfer of features from Arabic to English, such as the influence of phoneme-grapheme correspondence and phonological gaps. Grammatical errors were mostly attributed to interlingual interference, as learners heavily relied on Arabic sentence structures, though intralingual processes such as overgeneralization and simplification. This study highlights the importance of employing error analysis as an important tool to examine errors and emphasizes its effectiveness in identifying the linguistic areas of challenge for learners.

Following the same focus on grammatical issues, Donoso and Gómez (2018) explored the written discourse of future English teachers within the context of English as a foreign

language. The study employed a descriptive research design with participants from the first, second, third, and fourth years of a Chilean university, totaling 72 students. The research aimed to describe the interlanguage status of these learners and understand interlanguage development by examining their grammatical errors. The research posed a specific question regarding the most common grammatical errors made by these English teachers in training, considering their level of study and accounting for their interlanguage. A writing task was used, and a team of researchers, including an expert in English grammar, typified and quantified errors in the texts, focusing solely on grammatical errors. As expected, the findings revealed that fourth-year students made significantly fewer mistakes than first-year students. Also, the errors identified encompassed 86 types, with the most frequent being the incorrect use of prepositions and syntactic calques. The authors concluded that the development of interlanguage is related to students' level of study, with higher-level courses associated with fewer errors in written production. While the study did not explore possible causes of errors, the authors emphasized its usefulness in initiating academic discussions for future studies, particularly in understanding the potential fossilization of errors. This article is pertinent to our study as it illuminates the presentation of interlanguage in written discourse and provides a statistical analysis that can be helpful for our research.

From the review of the international literature on error analysis, it can be said that a significant portion of the studies is predominantly descriptive. For instance, numerous global studies within this research domain involve the identification, description, and categorization of errors present in students' written productions (Díez-Bedmar, 2011; Garrido & Rosado, 2012; Alawi, 2014; Donoso & Gómez, 2018). Researchers typically engage students in specific tasks or tests to investigate errors in their written compositions (Alawi, 2014; Donoso & Gómez, 2018).

For instance, scholars such as Garrido and Rosado (2012) adopted the approach with a different task, employing a translation of a letter and comparing the results of the written composition with an anticipated model. Their analysis focused solely on errors related to English tenses and aspects. Despite the utility of error analysis in extracting information about learners' language, it has notable limitations. Among the limitations that the authors mentioned, it can be seen that errors are highlighted rather than recognizing correct language use, an excessive emphasis on production data is placed, and the results are confined to specific populations, lacking generalizability.

2.2.1.2. Studies on socioeconomic conditions

The impact of socioeconomic conditions on EFL writing performance and language use is a key aspect in the present study. Many differences in the socioeconomic background are a variable worth considering due to the significant disparity that may exist among learners in EFL classrooms. Because of this, the following studies are considered as part of the empirical research for this study.

Firstly, Zare-ee (2010) investigated how learners' socioeconomic status (SES) influences English proficiency and academic writing among EFL undergraduates in Iran. The study aimed to determine if there is any significant difference in English proficiency and academic writing quality between students from high, average, and low SES backgrounds. In this study, the social constructivist perspective was used to emphasize the contextual nature of learning. Also, the research employed a quantitative, survey-based design. The participants were 196 third-year Iranian English-major undergraduates (48 male and 148 female) aged from 20 to 24. Data were gathered through a socioeconomic questionnaire assessing parental income, education, and

occupation: the Michigan English Language Assessment Battery (MELAB) for general proficiency, and an argumentative writing task rated by two experts using the ESL Composition Profile. The data obtained were analyzed using an analysis of variance and least significant difference (LSD) post hoc tests. Results revealed that students from higher SES backgrounds significantly outperformed their lower-SES peers in both English proficiency ($F=3.769$, $p=.025$) and academic writing quality ($F=3.632$, $p=.028$). Despite the uniform instruction and academic exposure, differences in socioeconomic status led to disparities in learners' literacy experiences and results, indicating that the advantages associated with higher SES persist into tertiary EFL learning. The study concluded that socioeconomic background shows sustained effects on EFL learners' academic writing and language proficiency, calling for greater pedagogical awareness and institutional support for underprivileged students.

In this second research in this regard, Adebileje, Adeleke, and Ajilore (2012) studied how socioeconomic factors influence the writing competence of junior secondary school students in Lagos, Nigeria. Based on sociolinguistic perspectives on education, this study used a survey design involving 300 Basic Nine students from 10 randomly selected public and private schools. Data were collected through a structured questionnaire that asked about parents' income, education, occupation, and language use at home, as well as through essay writing tests. The essays were scored using four main criteria: content, organization, expression, and mechanical accuracy (COEMA). Quantitative analysis using t-tests determined relationships between socioeconomic variables and writing competence. The findings revealed that around 66% of students were incompetent in writing, with public school students performing worse than students from private schools. Parental income and education were not significant factors, but the language spoken at home had a strong influence on writing ability. Because of this, the study

concluded that poor writing skills mainly result from weak grammar knowledge and limited use of English at home. It suggested that ESL teachers should focus more on grammar, punctuation, and writing practice, and that parents should take a more active role in supporting their children's education. For future research, this study recommended future research with larger and more diverse groups for longer periods.

Now, Dölek and Hamzadayi's (2018) research consisted of how socioeconomic status (SES) affects the writing skills of eighth-grade students in Gaziantep, Turkey. Guided by sociolinguistic theories, particularly Bernstein's deficit theory and Labov's theory of separation, the study explored the differences in writing performance between students from high and low SES backgrounds. This research employed quantitative methods, combining a survey design and a causal-comparative design. The participants were 120 eighth-grade students and 67 Turkish teachers from four schools in Gaziantep, selected through purposeful sampling based on the schools' socioeconomic environments during the 2016–2017 academic year. Data collection was carried out through students' written texts, a socioeconomic status survey, and teacher questionnaires evaluating students' writing processes. The written texts were assessed using a Written Expression Evaluation Form, and the data were analyzed using NCSS 2007 software, applying descriptive statistics and Mann-Whitney U tests. Results showed that students with higher socioeconomic status (SES) outperformed their lower-SES peers in all dimensions of writing (form, content formation and organization, word choice, and grammar). However, both groups shared similar difficulties in punctuation, structure, and idea development. Also, high-SES students demonstrated superior ability in evaluating their own writing. The study concluded that SES strongly influences writing performance, likely due to disparities in linguistic exposure and cognitive competence.

In summary, across the three studies, all researchers used quantitative methods to examine how socioeconomic status (SES) influences students' writing abilities, but they differed in scope, participants, and instruments. Zare-ee (2010) and Dölek and Hamzadayi (2018) focused on older EFL learners (university and eighth-grade students), while Adebileje et al. (2012) examined junior secondary students. Each study employed survey-based designs, with Dölek and Hamzadayi adding a causal-comparative element. Data collection typically included socioeconomic questionnaires (parental income, education, and occupation) and writing assessments, though the rating tools varied: Zare-ee used the ESL Composition Profile, Adebileje et al. used COEMA, and Dölek and Hamzadayi applied a Written Expression Evaluation Form. Lastly, findings consistently indicated that students from higher SES backgrounds performed better in writing, though the extent and causes differed. Zare-ee attributed disparities to sustained literacy advantages, while Dölek and Hamzadayi emphasized linguistic and cognitive exposure. In contrast, Adebileje et al. found home language use more influential than income or education.

2.2.1.3. Studies on writing enjoyment

Finally, another of the main focuses of this study is the enjoyment of writing by learners of English as a foreign language (EFL learners), given that this is one of the variables to be taken into account in the correlations carried out in the present study. Due to this, studies such as the following, which focus on aspects like motivation, attitudes, engagement, and writing enjoyment, are taken into account.

Setyowati and Sukmawan (2016) aimed to explore Indonesian EFL students' attitudes toward writing in English, their perceptions of the writing process, and the strategies they used to

improve their skills. The study adopted a descriptive quantitative design grounded in the theories of writing motivation, affect, and writing self-efficacy (Hidi & Boscolo, 2007; Pajares, Valiante, & Cheong, 2006). The study was conducted at STKIP PGRI Pasuruan, East Java, and it involved 57 students from an English Education program. Data were gathered through a writing attitude questionnaire, random interviews, and self-reflection essays, and analyzed using descriptive statistics and coding. Results showed that 52.63% of students had moderate attitudes and 47.37% positive attitudes toward writing. Students with moderate attitudes viewed writing as stressful and difficult due to grammar and vocabulary problems, while those with higher attitudes found it interesting and challenging. Common improvement strategies included practice writing (38.7%), reading extensively (26.3%), and diary writing (12.3%). The study concluded that positive attitudes encourage engagement, but confidence and feedback strongly influence performance. Limitations included its non-generalizable descriptive design and lack of correlation between attitude and writing performance. Future studies were suggested to examine attitude-performance relationships and proficiency level differences.

Continuing with this focus on attitudes towards writing, Raoofi and Maroofi (2017) conducted a study that examined the interrelationships among motivation (self-efficacy and task value), writing strategy use, and L2 writing performance among Malaysian undergraduate English second language (ESL) students. Grounded in social cognitive theory (Bandura, 1986) and expectancy-value theory, it explored how learners' beliefs about their own abilities and the value they assign to writing tasks influence strategy use and writing achievement. The sample they studied consisted of 304 students (ages from 19 to 24) enrolled in a general English writing course at a Malaysian public university. Data were collected through questionnaires measuring writing self-efficacy, task value, and strategy use, along with two writing performance tasks

(argumentative and descriptive essays) scored holistically following TOEFL guidelines.

Correlation and multiple regression analyzes revealed that all motivational constructs and most strategy categories were positively related to writing performance, though self-efficacy and intrinsic value emerged as the strongest predictors when other variables were controlled. Strategy use, while correlated with performance, lost its predictive power when motivation variables were included. The researchers concluded that learners' confidence in their writing abilities (self-efficacy) and interest in writing (intrinsic value) play crucial roles in L2 writing success.

They recommended that teachers nurture these factors through motivational and learner-centered pedagogies. The limitations of this study included the single-institution sample, cross-sectional design, and exclusion of other influential variables such as prior language knowledge or socioeconomic factors. Future research should adopt longitudinal designs, include diverse contexts, and investigate how motivational and cognitive factors interact over time to influence L2 writing development.

Additionally, another study directly related to this subject is the one carried out by Sabti et al. (2019), who conducted a quantitative correlational study to examine the relationships among writing anxiety, writing achievement motivation, writing self-efficacy, and writing performance among Iraqi tertiary EFL learners. Grounded on Bandura's social cognitive theory and research on motivation and affect in second language acquisition, the study aimed to understand how affective and motivational factors jointly influence students' writing achievement. The participants were 100 Iraqi undergraduate English majors from two public universities, all at an intermediate EFL proficiency level. Data were gathered through three standardized questionnaires (the Second Language Writing Anxiety Inventory (SLWAI), the Writer Self-Perception Scale (WSPS), and the Achievement Goal Questionnaire (AGQ)),

alongside a 200–250-word descriptive writing task evaluated through holistic scoring. Descriptive statistics, one-way analysis of variance, and Pearson correlation coefficient were used for analysis to determine the predictive power of the three variables on writing performance. The obtained results revealed that most students exhibited high writing anxiety, low achievement motivation, and low self-efficacy. Writing anxiety showed a significant negative effect on writing performance, whereas self-efficacy and achievement motivation correlated positively with each other and positively influenced writing outcomes. However, no significant relationship was found between anxiety and the other two factors. The study concluded that affective variables can significantly shape EFL writing performance, and recommended that teachers should cultivate students' motivation and confidence while minimizing anxiety through supportive feedback and collaborative writing practices. Some limitations included the exclusive focus on Iraqi undergraduates and the lack of gender analysis due to unbalanced samples. The authors recommend future research exploring gender differences and replicating the study with diverse EFL populations across educational levels.

Now, the following synthesis compares characteristics of these three studies, given their shared focus on affective and motivational influences in L2 writing and their distinct approaches to investigating these relationships. Across the three studies, Setyowati and Sukmawan (2016), Raoofi and Maroofi (2017), and Sabti et al. (2019), there is a shared focus on affective factors influencing EFL/ESL writing performance, though they differ in methodological rigor, instruments, and scope. Methodologically, all the studies employed quantitative approaches, but Setyowati and Sukmawan (2016) used a descriptive design, whereas the other two applied a correlational design to examine relationships among variables. In terms of data collection, the Indonesian study combined questionnaires, interviews, and self-reflection essays, integrating

qualitative insights into quantitative analysis. By contrast, Raoofi and Maroofi (2017) and Sabti et al. (2019) relied primarily on standardized questionnaires measuring self-efficacy, motivation, and related constructs, supplemented by writing performance tasks. Regarding findings, all three studies highlighted the pivotal role of affective and motivational factors in shaping writing performance. Setyowati and Sukmawan found that positive attitudes correlated with engagement but not directly with performance; Raoofi and Maroofi identified that self-efficacy and intrinsic task value are the strongest predictors of writing quality; and Sabti et al. confirmed that high anxiety can lower performance, while self-efficacy and motivation can enhance it. Their conclusions converge on the importance of confidence, motivation, and feedback in stimulating writing success, while limitations included context-specific samples, cross-sectional designs, and restricted generalizability, prompting calls for longitudinal and multi-context studies exploring affective and motivational dynamics in L2 writing.

2.2.2. National research

Expanding on the exploration of error analysis in language learning contexts, the following studies were carried out in Colombia. These studies extend their focus to include students' perceptions and sociodemographic factors, understanding that this interaction between language proficiency, learner attitudes, and demographic variables contributes to a holistic perspective on the challenges learners face in their written production processes.

In their study, García and Zuluaga (2021) conducted an exploratory analysis characterizing the written interlanguage of 21 intermediate-level students of Spanish as a foreign language at a private school called Spanish Adventure in San Carlos, Antioquia, Colombia. This study focused on analyzing the students' performance in written production processes with their

demographic and learning profiles. For this, the participants engaged in a communicative task of performance and completed a survey. Each text had an average of 538,9 words (11.298 words in total), and the error analysis revealed 1056 errors in the 21 texts, with the most frequent being omissions, particularly orthographic errors that did not significantly affect the message.

Regarding students' perceptions, they considered the difficulty of the test mostly normal and reported difficulties in grammar, lexical aspects, and strategic competence. The study also explored the students' perspectives on writing, revealing varied attitudes, ranging from considering writing as necessary to recognizing it as indispensable in their lives or work. The study concluded that not only the mother tongue but also the place of origin influenced students' performance in writing in Spanish, since the participants of this study came from North America and Europe. Due to its exploratory nature, the study abstained from making generalizations, emphasizing its focus on the action-educational context and its potential transferability to other contexts. The study also acknowledged the impact of the COVID-19 pandemic, the absence of local studies, and the limited knowledge of the students' repertoire.

Additionally, Pardo (2021) focused on analyzing written English texts produced by university students of EFL for the identification of errors that show their interlanguage. The main objective was to present a descriptive and classificatory analysis of the three most relevant errors in the students' written production. The participants were students from the Universidad del Norte, enrolled in EFL courses according to their performance levels. Data were collected during the second semester of 2015, utilizing surveys and a corpus. The error analysis and tagging followed the Error Tagging Manual Version 1.2 (Dagneaux et al., 2005) from the University of Louvain. In this study, the three most common errors were Spelling, Articles, and Individual Lexicon. The study concluded that grammatical aspects posed the most significant challenges for

EFL students, particularly in spelling. Errors in article usage suggested an ongoing change in students' interlanguage, persisting until level B2. Additionally, errors in individual lexicon were attributed to inappropriate vocabulary selection and incorrect collocation use. In this study, the authors acknowledged certain limitations, including the sample size, limited generalizability to a broader population, potential coding errors, and the need for further studies to ensure result comparability and draw more robust conclusions. Despite these limitations, the study's methodology utilizes an error tagging manual similar to the one being used in the present project, and it is framed within a university context, which makes it relevant to our project.

Following the same line of research, Pardo and Quiroz (2021) conducted a statistical analysis of spelling errors in a corpus of university-level students studying English as a foreign language. The objective was to determine if there is a relationship between the main errors in written compositions and two sociodemographic factors: gender and socioeconomic stratum. The analysis involved survey data and a corpus, with errors analyzed using the error analysis method and coded according to the error annotation system of the University of Louvain version 1.2. Although statistically non-significant, the results suggested that women produced fewer errors per 100 tokens, indicating a trend of better performance in learning English as a foreign language, consistent with previous studies. Additionally, socioeconomic status also showed unexpected results: while the differences did not reach statistical significance, the researchers found that students from lower strata made fewer mistakes than those from higher strata, challenging claims that individuals from wealthier backgrounds possess linguistic advantages in EFL written production provided by class privileges such as greater access to travel abroad and, therefore, an improvement in their level of English (Lin, 1999). The study acknowledged limitations related to partial participant information and the voluntary nature of data submission.

Despite these limitations, since this study used the same error analysis method as the ongoing research, it is particularly useful for comparison and as a guide to follow.

Another study of particular relevance is the one made by Chaparro (2022), a descriptive and qualitative study aimed to determine the role of Spanish L1 linguistic transfer in the written interlanguage of Colombian students learning German as a foreign language. This study was carried out with 12 intermediate-level (B1) students of German from the extension language course at the Universidad del Atlántico in Barranquilla, Colombia. The study methodology is based on analyzing a corpus of errors extracted from short narrative texts to identify, classify, and determine the cause of errors, particularly those resulting from negative L1 transfer. The results showed that negative transfer from Spanish significantly influenced students' written errors, occurring both directly (through the direct use of L1 structures) and indirectly (via literal translation, false analogy, or code-change).

2.2.3. Local research

Concluding our exploration of empirical research within the domains of interlanguage and error analysis, the following section focuses on three distinct studies conducted in Cali. These studies not only shed light on the complexities of English as a foreign language (EFL) learning but also provide important insights into diverse learner groups within the local community. By delving into these studies, we aim to foster a deeper understanding of the specific challenges and dynamics associated with EFL acquisition in Cali and identify gaps in the existing research landscape of the local context.

To start, despite the relevance of research on interlanguage and/or error analysis in the acquisition of English as a foreign language (EFL) by learners whose native tongue is Spanish,

not much research has been undertaken so far in the local context of Cali (Colombia) about this topic. Lemos (2022) used error analysis as a research method in language learning for English foreign written compositions of elementary school students. Also, Hoyos and Roldán (2016) carried out a study analyzing the syntactic errors that English learners committed due to the interference of Spanish as their mother tongue. Similarly, Chaves (2011) conducted mixed qualitative-quantitative research that focused on the use of error analysis as a tool for describing and explaining errors in written compositions of intermediate-level college students of English as a foreign language (EFL) at an undergraduate Foreign Languages program.

Lemos (2022) conducted an error analysis of the written productions of EFL elementary students at a bilingual private school to categorize the most common syntactical and vocabulary errors and provide pedagogical implications for EFL teachers and researchers who work with elementary students in bilingual contexts. For this, the general objective established was “to analyze different types of errors in written productions made by elementary students in a bilingual private school in Cali” (Lemos, 2022, p. 16). This study was designed following a descriptive approach and was conducted in a private bilingual school in Cali, Colombia. The participants were 144 elementary students from first to fifth grade whose performance was both very good and average/satisfactory, according to the institution’s method to grade performance. Besides, the data collected consisted of diverse written compositions and analyzed using a deductive approach based on the error analysis classification proposed by Richards (1974), who classified errors into two groups, interlingual (caused by L1 interference) and intralingual (caused by difficulties during the learning process). The results indicated that spelling and grammatical errors were the most common, particularly those related to verb tenses. Also, L1

interference was observed; however, the most frequent errors were intralingual and developmental.

In comparison, Hoyos and Roldán (2016) presented a study examining syntactic errors committed by English learners, caused by the influence of their native language, Spanish. The main objective that guided this research was to identify which are the most common errors in the order of words, pronouns, and prepositions, caused by the interference of the first language, in the composition of argumentative texts in English. The participants in this study were 43 students ranging between 19 and 25 years old, who belonged to the sixth semester of the undergraduate program in Foreign Languages at Universidad del Valle, in Cali, Colombia. Similar to previous studies, the analysis followed the four stages for error analysis proposed by Corder (1981), which are identification and classification, description, explanation and diagnosis of causes, and corrective proposals. The identified errors were classified based on linguistic/grammatical criteria, descriptive criteria, and etiological criteria. The linguistic/grammatical criteria focused on word order, pronoun usage, and preposition usage. The descriptive criteria determined if the errors involved omission, addition, false choice, and/or incorrect placement. The etiological criteria aimed to determine the origin of the errors, whether they were interlingual or intralingual, as Richards (1974) proposed. Hoyos and Roldán's (2016) study found that observed errors concerned the use of prepositions, word order, and pronouns. Through the survey, the researchers also found that most students acknowledged the need for improvement in their English writing skills, as well as they recognized the noticeable progress they had made through the program.

Overall, the reviewed literature on error analysis in EFL writing reveals very interesting information about this field, diverse approaches, and findings. Firstly, international studies offer a strong foundation, highlighting the prevalence of descriptive research exploring common error types in written compositions. Studies by Díez-Bedmar (2011), Garrido and Rosado (2012), Alawi (2014) and Donoso and Gómez (2018) exemplify this trend, identifying areas like grammar, verb usage, and punctuation as frequent sources of errors. However, these studies often lack in-depth analysis of error causes and neglect factors such as student perceptions and demographics.

Secondly, some national researches expand upon this foundation by considering student perspectives and sociodemographic influences. For instance, García and Zuluaga (2021) and Pardo and Quiroz (2021) illustrate this broadening of research by exploring the interplay between language proficiency, learner attitudes, and demographic variables. Notably, Pardo (2021) provides valuable insights relevant to our project, employing the same error tagging manual and focusing on university-level EFL contexts.

Lastly, local research within Cali offers a crucial lens for understanding the specific challenges and dynamics of EFL acquisition in our context. Lemos (2022), Hoyos and Roldán (2016), and Chaves (2011) shed light on this landscape, examining error types, L1 interference, and the use of error analysis as a pedagogical tool. Moreover, it is noteworthy that their results underscore the need for further research addressing knowledge gaps specific to Cali's EFL learners, thereby supporting and justifying the necessity for research studies akin to the present one in this context.

In general, the reviewed literature provides a comprehensive framework for our study. By building upon international approaches, incorporating insights from national research, and addressing specific gaps identified in the local context, our study can contribute significantly to the understanding of EFL writing through error analysis of pre-service teachers at Universidad del Valle.

RESEARCH METHODOLOGY

This section provides a detailed description of the research design and the methodology used in this study. First, it describes the research approach and design according to the purpose of the study. Secondly, it describes the characteristics of the population and the place under study. Thirdly, a description of the data collection, procedures and analysis methods. Finally, it presents the principles of validity and reliability used in the study.

3.1. Research approach and design

Since the main purpose of this study is to find the most common errors in the written interlanguage of a group of students and explore if there is a correlation between the errors and individual factors, this study is situated within the quantitative paradigm. As defined by Creswell (2012), this approach involves delineating a research problem by elucidating trends or explaining the necessity for an exploration of the connections between variables. According to Creswell (2012), the extent of correlation in this type of research signifies whether there is a connection between variables or whether one variable can be used to predict another. Hence, the design of this research is correlational, emphasizing the examination of variables such as social strata and writing enjoyment.

Taking into account that the study focuses on the most common errors in the written interlanguage (a unique transitional stage between native language and target language), the methodology used is based on the Computer-Aided Error Analysis (CEA) approach, which facilitates the identification and classification of learner errors through the use of learner corpora and specialized digital tools (Granger, 2008). Learner corpora, as defined by Granger (2012) and

Bennett (2010), are a type of specialized corpus composed of electronic collections of texts (either written or transcribed speech) produced by individuals acquiring a second or foreign language. The broader field in which this approach is situated is Corpus linguistics, which involves the study of language through the exploration of large, electronic collections of texts that have annotations for the systematic investigation of language patterns (Kennedy, 1998). This methodology allows researchers to investigate recurrent features and discrepancies in learner language across contexts.

3.2. Research site and participants

3.2.1. Research site

This research takes place in the Bachelor's Degree in Foreign Languages at the School of Language Sciences of the Universidad del Valle in Cali, Colombia. The main objective of this undergraduate program is to train graduates in foreign languages with an emphasis on English and French, who can teach these two languages in secondary education. The mission of this program is to study language as a sociocultural phenomenon, a communication system, and a means of construction and expression of identity and knowledge. Additionally, the program aims to train professionals in foreign languages who are capable of working in teaching and research with a social projection.

The curricular components of the program are pedagogy and didactics, research training, Linguistic foundation, and foreign languages: English and French. The training in these two foreign languages is structured in three phases: the development of basic communication skills, the deepening of the study of oral and written texts, and the literature of English and French. In

this context, writing is a fundamental skill that students develop throughout their degree, developing this ability through the creation of various types of texts, such as essays and reviews, as well as by analyzing information, including statistical charts, to provide well-informed insights.

3.2.2. *Participants*

The participants in this study consisted of 49 eighth-semester students enrolled in the Bachelor's Degree in Foreign Languages at the School of Languages Sciences of the Universidad del Valle in Cali, Colombia. These students, whose first language is Spanish, are required to demonstrate proficiency at a C1 level (as established by the CEFR) as part of the graduation requirements of the program. At this proficiency level, language learners are expected to produce clear and detailed written discourse on a wide range of subjects, including abstract topics, while demonstrating a degree of precision and nuance in their language use (Council of Europe, 2017). Although the corpus for this study comprises the writing samples of 49 students, only 30 voluntarily completed a survey designed to collect information related to their sociodemographic and personal information. This survey revealed a diverse age range among participants: approximately 63% were between 20 and 24 years old, 16% between 25 and 30, and 20% were over 30 years of age.

It must be noted that this cohort is particularly relevant given their advanced stage in the curriculum, which includes several and intense courses in language skills and pedagogical training. At this point, they are expected to demonstrate their knowledge and also their ability to apply it in academic and teaching contexts. Therefore, analyzing their written production in

relation to other variables offers valuable perspectives for both linguistic research and teacher education.

The following section outlines the methodology used to analyze their written productions and explore how variables are related to their English Language development.

3.3. Data collection methods, procedures and analysis

To examine learners' written performance and the influence of individual and sociocultural factors, two primary sources of data were utilized: written compositions produced during an English proficiency test and a virtual survey completed by participants. The methodology includes the design of the writing tasks, the conditions under which the data was collected, the process of error tagging, and the statistical tools employed for data analysis.

3.3.1. *Data collection methods*

3.3.1.1. Students' written compositions

The corpus used for this research consisted of 49 written argumentative compositions authored by 49 students of the undergraduate Degree in Foreign Languages program of the Universidad del Valle, written for an English proficiency test. This assessment evaluates English language competency across all skills and is administered when they finish the fourth year of the academic program. According to the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001), at this level, the students should be able to produce clear and detailed texts on a wide range of topics and explain a point of view on a discussion topic by presenting the advantages and disadvantages of various options.

These texts were produced under exam conditions for the English proficiency test destined for intermediate-level language learners, culminating their 8th semester in the academic program during the COVID-19 pandemic, a timeframe characterized by lockdowns and biosafety protocols. Consequently, students presented their proficiency examinations in a virtual modality. To facilitate this process, the examination was administered through Campus Virtual, the online Moodle platform of the Universidad del Valle. To uphold the exam conditions during the English proficiency test, students were not allowed to use other programs or websites while using Campus Virtual. Moreover, they were deprived of access to dictionaries, other electronic devices, and tools for orthographic and grammatical review and correction during the writing section of the examination.

For the writing section of the proficiency test, the students were required to undertake two writing assignments of 250 words each within a timeframe of 90 minutes, corresponding to 45 minutes for each task. The first task consisted of the composition of an opinion essay, and the second task entailed the composition of a text based on the interpretation and explanation of data from a graph. Figure 1 illustrates the prompt used in this kind of test. It is essential to mention that this was a one-time writing assignment, with no feedback given by teachers and without the use of any kind of tool or spell-checker.

Figure 1*Example of a Trial Test*


Writing trial test

00-LLE-EXAMEN-COMPETENCIAS-INGLES > Writing trial test



This section of the test measures your ability to demonstrate cognitive abilities (planning, writing, monitoring, and reviewing) and digital skills (typing skills, computer familiarity) to produce a text. It also measures your ability to produce non-specialized academic texts addressed to an academic audience within a limited time. You are required to produce structured and coherent texts that reflect a good use and variety of vocabulary, register, punctuation, spelling, and grammatical structures.

In this section, you will perform two writing tasks and are given 90 minutes to complete them (45 mins each).

For the first writing task, you are required to produce an academic essay that states, explains, and develops your position on a familiar topic, providing specific details and examples.

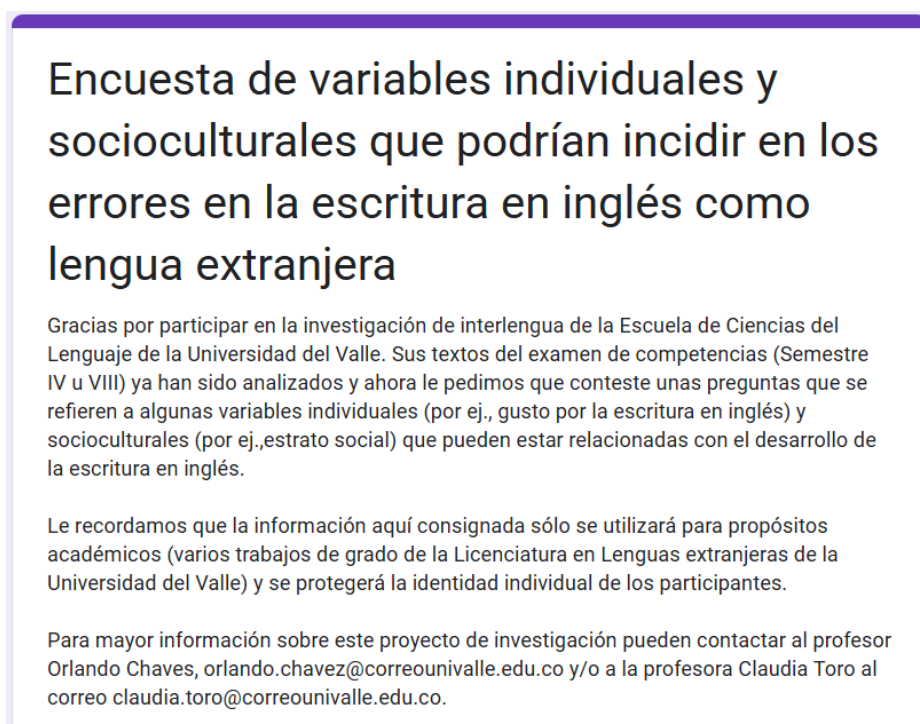
For the second writing task, you are required to interpret, present clearly and correctly and comment on the most important points or trends contained in diagrams or other visual data on familiar or professional topics.

Each text will be assessed by two independent raters using an analytic scoring rubric with a scale from 0 to 5. The two scores will be averaged to obtain the final score.

Remember that this trial test is similar in structure to the real test.

3.3.1.2. Survey

For this research, a structured survey was conducted online through Google Forms, utilizing a closed-ended question format to collect essential information about the individual and sociocultural particularities of the participants, ensuring an optimal understanding of their particular contexts. To maximize clarity and comprehension, this survey was administered in Spanish, the first language of the participants, as shown in Figure 2. Also, the survey was strategically segmented into two primary sections to collect the pertinent data on their individual and sociocultural variables.

Figure 2*Title and Introduction of the Survey*


Encuesta de variables individuales y socioculturales que podrían incidir en los errores en la escritura en inglés como lengua extranjera

Gracias por participar en la investigación de interlengua de la Escuela de Ciencias del Lenguaje de la Universidad del Valle. Sus textos del examen de competencias (Semestre IV u VIII) ya han sido analizados y ahora le pedimos que conteste unas preguntas que se refieren a algunas variables individuales (por ej., gusto por la escritura en inglés) y socioculturales (por ej., estrato social) que pueden estar relacionadas con el desarrollo de la escritura en inglés.

Le recordamos que la información aquí consignada sólo se utilizará para propósitos académicos (varios trabajos de grado de la Licenciatura en Lenguas extranjeras de la Universidad del Valle) y se protegerá la identidad individual de los participantes.

Para mayor información sobre este proyecto de investigación pueden contactar al profesor Orlando Chavez, orlando.chavez@correounivalle.edu.co y/o a la profesora Claudia Toro al correo claudia.toro@correounivalle.edu.co.

The first section, ‘Factores individuales’ (individual factors), was composed of eight questions where the participants were consulted about different aspects of their demographics and language backgrounds. These aspects were age, sex, mother tongue, liked aspects of the English language (e.g., writing, reading, grammar, and culture), other languages studied or spoken, predominant language in daily used digital devices, and English-speaking countries of residence and study. Although all the questions in this section were important for understanding potential variables influencing students’ written production, only the one regarding the participants’ liked aspects of the language (Figure 3) was selected for further analysis, as it provides insight into students’ individual interests that may have a more direct impact on their performance in writing tasks.

Figure 3*Question on Liked Aspects of the English Language*

4. Califíque su gusto por los siguientes aspectos de la lengua inglesa *
(1 -muy poco- a 5 -muchísimo-)

	1 (Muy poco)	2 (Poco)	3 (Neutral)	4 (Mucho)	5 (Muchísimo)
El idioma inglés, en general	☐	☐	☐	☐	☐
La escritura en inglés	☐	☐	☐	☐	☐
La lectura en inglés	☐	☐	☐	☐	☐
La gramática del inglés	☐	☐	☐	☐	☐
La cultura de los países angloparlantes	☐	☐	☐	☐	☐

The second section, ‘Factores socioculturales’ (sociocultural factors), was integrated by four more questions that aimed to explore particular aspects of their socioeconomic and cultural backgrounds. These questions asked the participants about their socioeconomic stratum, if they studied at a bilingual school or language institute, and the contexts out of the classroom where they regularly used English (e.g., on social media, with friends, with relatives, for work, etc.). Similar to the previous section, even if the additional information is useful to better understand the contextual circumstances of each participant, this research is specifically focused on the socioeconomic stratum influence on students’ written performance for deeper analysis (Figure 4).

Figure 4

Question on the Socioeconomic Stratum of the Participants

The image shows a screenshot of a survey interface. At the top, there is a purple header bar with the text 'Factores socioculturales' in white. Below this, the question '9. Estrato socio-económico *' is displayed. Underneath the question, there are six radio button options listed vertically: 'Estrato 1 (bajo bajo)', 'Estrato 2 (bajo)', 'Estrato 3 (medio-bajo)', 'Estrato 4 (medio)', 'Estrato 5 (medio-alto)', and 'Opción 6 (alto)'. Each option is preceded by an empty radio button circle.

To better understand the Colombian measurement of socioeconomic stratum, Colombia's National Administrative Department of Statistics (DANE) considers six hierarchical tiers (called 'estratos') that are designed to classify residential properties according to their physical infrastructure and neighborhood environment, rather than by household economic income. The classification goes as follows: stratum 1 (lowest-income class), stratum 2 (low-income class), stratum 3 (lower-middle class), stratum 4 (middle class), stratum 5 (upper-middle class), and stratum 6 (wealthiest class). The primary goal is to cross-subsidize public utilities like electricity, water, and gas, ensuring that wealthier households help to compensate costs of lower-income ones. In this manner, strata 1, 2, and 3 correspond to low-income tiers of households with limited economic resources, who are eligible for subsidies on residential public utilities. In contrast, strata 5 and 6 represent high-income tiers accommodating households with the greatest financial means, who are required to pay surcharges (contributions) on the cost of residential public

services. Stratum 4 does not qualify for subsidies nor is subject to surcharges; instead, households in this tier pay the amount defined by the service provider as the base cost of the utility (DANE, 2021).

3.3.2. *Procedures and analysis*

3.3.2.1. Error analysis

To perform the analysis of the data, the texts resulting from the proficiency test were converted into a plain text format, and errors were consistently tagged following the taxonomy and tagging procedures in the Error Tagging Manual from the University of Louvain (Dagneaux et al., 2005). This manual provides a detailed framework for tagging linguistic errors in non-native writing, based on a corpus-informed classification system. It defines eight main error categories (F: form, G: grammar, L: lexis, Q: punctuation, S: style, W: missing words, redundant words and word order, X: lexico-grammar, Z: infelicities), each one with specific subcategories. A total of 56 error tags are used, with each tag beginning with a letter that indicates the general category of the error (e.g., G for grammar, L for lexis), followed by additional characters to specify the subcategory and type of error. The goal of this system is to ensure consistent and precise tagging, even in cases involving overlapping or ambiguous errors. The full error categorization system and their abbreviations can be seen in Table 1.

Table 1*Error Categories and Their Abbreviations.*

Abbreviation	Category Description
FM	Form, Morphology
FS	Form, Spelling
FSR	Form, Spelling, Regional
GDD	Grammar, Determiner, Demonstrative
GDO	Grammar, Determiner, Possessive
GDI	Grammar, Determiner, Indefinite
GDT	Grammar, Determiner, Other
GA	Grammar, Articles
GADJCS	Grammar, Adjectives, Comparative / Superlative
GADJN	Grammar, Adjectives, Number
GADJO	Grammar, Adjectives, Order
GADVO	Grammar, Adverbs, Order
GNC	Grammar, Nouns, Case
GNN	Grammar, Nouns, Number
GPD	Grammar, Pronouns, Demonstrative
GPP	Grammar, Pronoun, Personal
GPO	Grammar, Pronoun, Possessive
GPI	Grammar, Pronoun, Indefinite
GPF	Grammar, Pronoun, Reflexive/Reciprocal
GPR	Grammar, Pronoun, Relative/ Interrogative

GPU	Grammar, Pronoun, Unclear reference
GVAUX	Grammar, Verbs, Auxiliaries
GVM	Grammar, Verbs, Morphology
GVN	Grammar, Verbs, Number
GVNF	Grammar, Verbs, Non-Finite / Finite
GVT	Grammar, Verbs, Tense
GVV	Grammar, Verbs, Voice
GWC	Grammar, Word Class
LCC	Lexis, Conjunctions, Coordinating
LCLC	Lexis, Connectors, Logical, Complex
LCLS	Lexis, Connectors, Logical, Single
LCS	Lexis, Conjunctions, Subordinating
LP	Lexical Phrase
LPF	Lexical Phrase, False friends
LS	Lexical Single
LSF	Lexical Single, False friends
QC	Punctuation, Confusion
QL	Punctuation, Lexical
QM	Punctuation, Missing
QR	Punctuation, Redundant
SI	Sentence, Incomplete
SU	Sentence, Unclear
WM	Word Missing
WO	Word Order

WRS	Word Redundant Single
WRM	Word Redundant Multiple
XADJCO	Lexico-Grammar, Adjectives, Complementation
XADJPR	Lexico-Grammar, Adjectives, Dependent Preposition
XCONJCO	Lexico-Grammar, Conjunctions, Complementation
XNCO	Lexico-Grammar, Nouns, Complementation
XNPR	Lexico-Grammar, Nouns, Dependent Preposition
XNUC	Lexico-Grammar, Nouns, Uncountable / Countable
XPRCO	Lexico-Grammar, Prepositions, Complementation
XVCO	Lexico-Grammar, Verbs, Complementation
XVPR	Lexico-Grammar, Verbs, Dependent Preposition
Z	Infelicities.

After establishing the taxonomy of errors presented in Table 1, the 49 texts resulting from the proficiency test were converted into plain text format and annotated, as illustrated in Figure 5. The tagging process involved placing the error tag in parentheses before the erroneous word or phrase, followed by the corrections made in dollar signs. This annotation facilitated further quantitative and qualitative analysis of the frequency and distribution of errors made by students across the corpus.

Figure 5*Example of Error Tagging in a Plain Text Format.*

Local, always the best option

(FS) Regarding \$Regarding\$ the fact that we are near to (GA) 0 \$the\$ holidays season, we must (XVPR) take in account \$take into account\$ that people use this time to travel and visit new places. However, there (GVN) have \$has\$ to exist a way in (FS) wich \$which\$ people get a more significant experience from these trips and learn more about (GPP) it \$them\$. From the way I see things, the best option to get a closer look (LS) of \$at\$ places is exploring through local residents' eyes, in this sense you will have a (GADJCS) much clear \$clearer\$ and (GVM) contextualize \$contextualized\$ background of (GNN) subject \$subjects\$ such as food, history, politics, economy, culture and social issues from the city you are visiting.

As (GVM) mention \$mentioned\$ before, going to a place is not only staying at a hotel or (FS) appartament \$apartment\$ or just (FS) swimming \$swimming\$ on the beach, traveling is also related (LS) with \$to\$ the way (GNN) tourist \$tourists\$ learn and interact with the whole city. Of course (QM) 0 \$,\$ going to a museum can (FS) aknowledge \$acknowledge\$ people (FS) fom \$from\$ historical background and a guided tour could mention some (SU) important accomplishment \$\$ that (FS) occured \$occurred\$ some time ago but I (Z) don't \$do not\$ see this as (GA) 0 \$a\$ significant experience. (GNN) Tourist \$Tourists\$ can hear this information and forget it the next day because they (GVT) found \$find\$ this as a lecture and not as (GA) a \$an\$ important (FS) experince \$experience\$. Hearing the history by local residents is meaningful since (WM) they not \$they do not\$ tell an isolated story, but they tell their life experience and how things impacted them.

3.3.2.2. Text alignment

After the tagging of the texts, all files were aligned by using the Concord tool from the software WordSmith Tools v5.0.0.334 (Scott, M., 2008) to count the number of errors per subcategory and organize the data for analysis. For this, it was necessary to first prepare a .txt file with the list of all the 56 subcategory tags (the ones from the Error Tagging Manual). This file was imported into Concord as the list of search words, working as the reference the software would follow to identify, sort and count each occurrence of the errors contained in the used corpus (as displayed in Figure 5). In this way, it was also necessary to individually process with Concord each of the 49 texts, to, as mentioned, extract information on the number of identified

errors in each particular text, with the ensured systematic and precise analysis provided by the software.

Figure 6

Example of Text Alignment Process in Concord Tool

Concord

FileEditViewComputeSettingsWindowsHelp

N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	lead	lead	Sec	Sec	File	%
1	in some cases more beautiful but less (FM) known \$known\$, we (GVT) lost	FM		330	430%		075%				075%	59ITC-8EOP2.tx	75%	
2	\$especially\$ if he/she is a close friend. (FS) Dispite \$Despite\$ (LCLS) Dispite	(FS)		183	3	4%	041%				041%	59ITC-8EOP2.tx	39%	
3	, details about the language (QL), \$or\$ (FS) expresions \$expressions\$ we	(FS)		224	378%		051%				051%	59ITC-8EOP2.tx	49%	
4	season \$seasons\$ to travel, which (FS) recomendations	(FS)		114	173%		026%				026%	59ITC-8EOP2.tx	24%	
5	to understand a native conversation. (FS) furthermore \$Furthermore\$, when	(FS)		237	4	2%	054%				054%	59ITC-8EOP2.tx	53%	
6	hidden places, the most delicious (FS) typical \$typical\$ food (QM) 0 \$,\$	(FS)		416	575%		034%				034%	59ITC-8EOP2.tx	94%	
7	places that are chosen to be visited by (FS) turist \$tourist\$ (GNN) turist	(FS)		310	471%		070%				070%	59ITC-8EOP2.tx	70%	
8	role and is a decisive issue because (FS) occasionally \$occasionally\$ (GPP)	(FS)		262	425%		059%				059%	59ITC-8EOP2.tx	59%	
9	resident may not (WRS) to \$offer\$ (GA) a \$certified\$ guidance, he or she	(GA)		198	331%		045%				045%	59ITC-8EOP2.tx	43%	
10	aspects of it, we can firstly search on (GA) 0 \$the\$ internet facts and reviews	(GA)		76	138%		017%				017%	59ITC-8EOP2.tx	16%	
11	, and some of them also like to learn in (GA) a \$the\$ deepest way the history,	(GA)		22	053%		0	5%			0	5%	59ITC-8EOP2.tx	5%
12	aspects we (Z) can't \$cannot\$ find on (GA) 0 \$the\$ internet or in a guided tour.	(GA)		433	532%		038%				038%	59ITC-8EOP2.tx	98%	
13	sometimes falls shorts due to the fact (GDD) 0 \$that\$ in this kind of options	(GDD)		291	453%		056%				056%	59ITC-8EOP2.tx	67%	
14	can show us (WR) indeed \$without\$ (GDI) 0 \$any\$ economic interest the	(GDI)		403	532%		031%				031%	59ITC-8EOP2.tx	91%	
15	to be visited by (FS) turist \$tourist\$ (GNN) turist \$tourists\$ and with some	(GNN)		313	474%		071%				071%	59ITC-8EOP2.tx	71%	
16	we can use to know which are the best (GNN) season \$seasons\$ to travel,	(GNN)		108	157%		024%				024%	59ITC-8EOP2.tx	22%	
17	(FS) occasionally \$occasionally\$ (GPP) there \$they\$ are cheap and	(GPP)		265	428%		050%				050%	59ITC-8EOP2.tx	60%	
18	(GPP) there \$they\$ are cheap and (GVN) includes \$include\$ food,	(GVN)		271	434%		051%				051%	59ITC-8EOP2.tx	61%	
19	Most people enjoy (GVNF) to know \$knowing\$ new places	(GVNF)		3	0	9%	0	1%			0	1%	59ITC-8EOP2.tx	1%
20	but less (FM) known \$known\$, we (GVT) lost \$lose\$ the opportunity to	(GVT)		334	433%		076%				076%	59ITC-8EOP2.tx	76%	
21	place \$the\$ (QM) 0 \$,\$ etc. When we (GVT) finally decide \$have\$ finally	(GVT)		147	210%		033%				033%	59ITC-8EOP2.tx	31%	
22	is a close friend. (FS) Dispite \$Despite\$ (LCLS) Dispite \$Although\$ a local	(LCLS)		186	3	9%	042%				042%	59ITC-8EOP2.tx	40%	
23	to know it is talking to a local resident. (LS) specially \$specially\$ if he/she is	(LS)		174	279%		039%				039%	59ITC-8EOP2.tx	37%	

concordance

collocates

plot

patterns

clusters

filenames

follow up

source text

notes

36SetFilename

In order to compile and categorize the data extracted from the individual analyzes carried out with Concord Tools, a Google Sheet was used, as shown in Figure 7. This Google Sheet contains, firstly, information about characteristics of the original compositions (such as the number of words, paragraphs and sentences), secondly, the data resulting from the error tagging and alignment (number of errors per type, per category, per text and the total of errors, as well as their corresponding average), and finally, the data collected from the survey on individual and sociocultural factors.

Figure 7*Categorized Data on Texts and Found Errors in Google Sheets*

highlighted in green: with survey (x30)

Participant/Text (ID code) (Green: with survey, x29)	#Words	#Paragraphs	#Sentences	ADJPR	(XCONJCO)	(XNCO)	(XNPR)	(XNUC)	(XPRCO)	(XVCO)	(XVPR)	(Z)	TOTAL ERRORS (out of 56)
074UCFP-8EOp3	500	5	24	0	0	0	0	0	0	0	1	1	21
076SRP-8EOp3	359	5	13	0	0	1	0	0	0	1	0	2	35
078DFOP-8EOp3	473	4	19	0	0	0	0	0	0	0	0	2	54
080LMGS-8EOp3	254	4	11	0	0	0	0	0	0	0	0	1	10
081MMBB-8EOp3	274	4	13	0	0	0	1	0	0	1	1	0	21
084JSM-8EOp4	282	4	7	0	0	0	1	0	0	0	0	0	50
055CCV-8EOp2	396	4	10	0	0	0	0	0	0	0	0	3	13
TOTAL (SUMA)	11948	136	430	1	0	1	7	0	7	6	13	50	799
TOTAL (PROMEDIO)	398	5	14	0,0	0,0	0,0	0,2	0,0	0,2	0,2	0,4	1,7	26,6
085AFGR-8EOp4	447	5	21	0	0	0	0	0	0	1	3	2	41
086SSJMM-8EOp4	267	3	10	0	0	0	0	0	0	0	2	2	33
087MFVP-8EOp4	342	3	16	0	0	0	0	0	0	0	0	0	31
043AMSC-8EOp1	239	4	11	0	0	0	0	0	0	0	1	1	11
046YETR-8EOp1	402	4	20	0	0	0	0	0	0	0	0	0	8
047DML-8EOp1	468	4	22	0	0	0	0	0	0	0	2	1	32
048ASPG-8EOp1	497	4	12	0	0	0	0	0	0	0	0	0	11
049SCZ-8EOp1	444	4	22	1	0	0	1	0	0	0	0	0	35
050JASR-8EOp1	517	5	24	0	0	0	0	0	1	0	0	3	38
056DFPI-8EOp2	459	6	22	0	0	0	0	0	0	0	2	0	24
060VCG-8EOp2	284	6	7	0	0	0	2	1	0	0	0	1	15
062VSTA-8EOp2	278	4	12	0	0	0	0	1	0	0	0	0	22
067NVMM-8EOp2	340	5	10	0	0	0	0	0	0	0	0	0	11
071LVMP-8EOp3	293	7	7	0	0	0	0	0	0	0	0	1	24
075DAPB-8EOp3	646	6	16	0	0	0	0	0	0	2	0	13	92
077GPSF-8EOp3	204	4	6	0	0	0	0	0	0	0	0	0	23
079SANG-8EOp3	341	4	16	0	0	0	0	0	0	0	0	0	1
082ELMP-8EOp3	354	4	8	0	0	0	0	0	1	1	0	0	31
083AMDC-8EOp3	367	5	9	0	0	0	1	0	0	1	1	0	87
TOTAL (SUMA)	19137	223	701	2	0	1	11	2	9	11	24	74	1369
TOTAL (PROMEDIO)	391	5	14	0,04	0,00	0,02	0,22	0,04	0,18	0,22	0,49	1,51	27,94

3.3.2.3. Peer review

As part of the data analysis process and a validation and reliability strategy, a peer review was conducted by an expert professor, who was responsible for evaluating the tagging of each of the 1,368 errors found in the 49 written compositions during the error analysis process, and providing fundamental feedback on the necessary corrections.

This review was conducted immediately after the alignment of the individual texts, also employing the Concord tool from the software package WordSmith Tools (Scott, M., 2008). The procedure to obtain the aligned data for the review consisted of following a process similar to

that used for the alignment of individual texts, as previously described. Specifically, it involved importing the list of search words (error tags) that the program would use for the alignment, as well as importing all 49 previously tagged texts into Concord as the files to be aligned.

After the precise analysis was provided by Concord, the resulting information was extracted from the software by saving it as an Excel spreadsheet containing the total of the 1,368 aligned errors from the 49 written essays. All of these errors were reviewed by the professor in charge, who provided comments and feedback on aspects such as incorrect tagging, procedural errors, and other necessary clarifications.

Figure 8

Alignment of Total Errors for Professor's Review

N	Concordance	Set	Tag	File
1	(GVN) engloves \$englobe\$ our (FM) socioeconomical \$socioeco	(FM)		040DFSG-8EOP1.txt
2	t they fail it or they do not (FM) completed \$complete\$ it.	(FM)		045JRV-8EOP1.txt
3	, these are necessary to keep (FM) inform \$informed\$ or to r	(FM)		044CRG-8EOP1.txt
4	ractice \$practices\$ of use of (FM) technologic \$technologica	(FM)		080LMGS-8EOP3.txt
5	\$are not\$ a huge part of our (FM) lifes \$lives\$ is useless,	(FM)		075DAPB-8EOP3.txt
6	mind that they are in a very (FM) delicated \$delicate\$ age,	(FM)		075DAPB-8EOP3.txt
7	s that they will have if they (FM) failed \$fail\$ (WRM) the t	(FM)		045JRV-8EOP1.txt
8	a bigger part of what a place (FM) consideres \$considers\$ as	(FM)		064DXQM-8EOP2.txt
9	ut whether students should be (FM) require \$required\$ to tak	(FM)		044CRG-8EOP1.txt
10	only fail the test but to be (FM) enabled \$unable\$ to pass	(FM)		043AMSC-8EOP1.txt
11	cases more beautiful but less (FM) knowed \$known\$, we (GVT)	(FM)		069ITC-8EOP2.txt
12	e (GNC) studentss \$students'\$ (FM) proeficiency \$proficiency	(FM)		049SCZ-8EOP1.txt
13	RS) in \$0\$ part of children's (FM) lifes \$lives\$.	(FM)		083AMDC-8EOP3.txt
14	s who do not agree with being (FM) evaluate \$evaluated\$ in a	(FM)		044CRG-8EOP1.txt
15	\$, affecting aspects of their (FM) lifes \$lives\$ such as rea	(FM)		081MMBB-8EOP3.txt
16	a, Cali, etc. Travelers often (FM) chose \$choose\$ challengin	(FM)		061DMMZ-8EOP2.txt
17	ey are in. On the other hand, (FM) taking \$talking\$ to a loc	(FM)		062VSTA-8EOP2.txt
18	e book is able to impact many (FM) life's \$lives\$, what can	(FM)		087MFVP-8EOP4.txt
19	ts is (GA) a \$an\$ alternative (FM) than \$that\$ can have both	(FM)		047DML-8EOP1.txt
20	n how that new culture works, (FS) nowadays \$nowadays\$ in ord	(FS)		062VSTA-8EOP2.txt
21	sult, they will gain a deeper (FS) undstanding \$understandi	(FS)		062VSTA-8EOP2.txt
22	tal health problems such as: (FS) Attention \$attention\$ pro	(FS)		082ELMP-8EOP3.txt
23	der to protect the health and (FS) well being \$well-being\$ o	(FS)		082ELMP-8EOP3.txt
24	you can join a guided tour in (FS) wich \$which\$ you can star	(FS)		066JJMS-8EOP2.txt
25	\$ understand it as (GA) 0 \$a\$ (FS) momentanous \$momentaneous	(FS)		047DML-8EOP1.txt
26	orm \$from\$ the learners, that (FS) wil \$will\$ understand it	(FS)		047DML-8EOP1.txt
27	stiy, (GWC) thought \$through\$ (FS) joing \$joining\$ a guided	(FS)		062VSTA-8EOP2.txt
28	as kids they should have some (FS) guidance \$guidance\$ from	(FS)		082ELMP-8EOP3.txt
29	hysical health problems like: (FS) Early \$early\$ back proble	(FS)		082ELMP-8EOP3.txt
30	electronic devices. Nowadays, (FS) technologie \$technology\$	(FS)		082ELMP-8EOP3.txt
31	the city or the place works. (FS) on \$On\$ one hand, people	(FS)		062VSTA-8EOP2.txt
32	omputers and other electronic (FS) deviceess \$devices\$ since	(FS)		082ELMP-8EOP3.txt

3.3.2.4. Correlations

To establish the correlations, the categorized data obtained from the error tagging was analyzed using the software JASP 0.18.1 (JASP Team, 2023) to find out the descriptive statistics and correlations between the number of errors and individual variables, specifically socioeconomic stratum and writing enjoyment. The correlational analysis consisted of three main points: the correlation between the total number of errors and students' socioeconomic strata, the correlation between the total number of errors and students' writing enjoyment, and the correlation between students' writing enjoyment and socioeconomic strata.

To address this part of the process, some information from the Google Sheet shown in Figure 8 was used to create another spreadsheet containing the required data to be used in the software JASP to carry out the analyses on descriptive statistics and correlations. The information selected to create this new file consisted of the text identification (*ID*), the number of errors per text (*ERRORS*), and the responses obtained from the survey about the students' writing enjoyment (*LikeWriting*) and socioeconomic stratum (*SOCIAL-STRATA*). The resulting table can be seen in Figure 9.

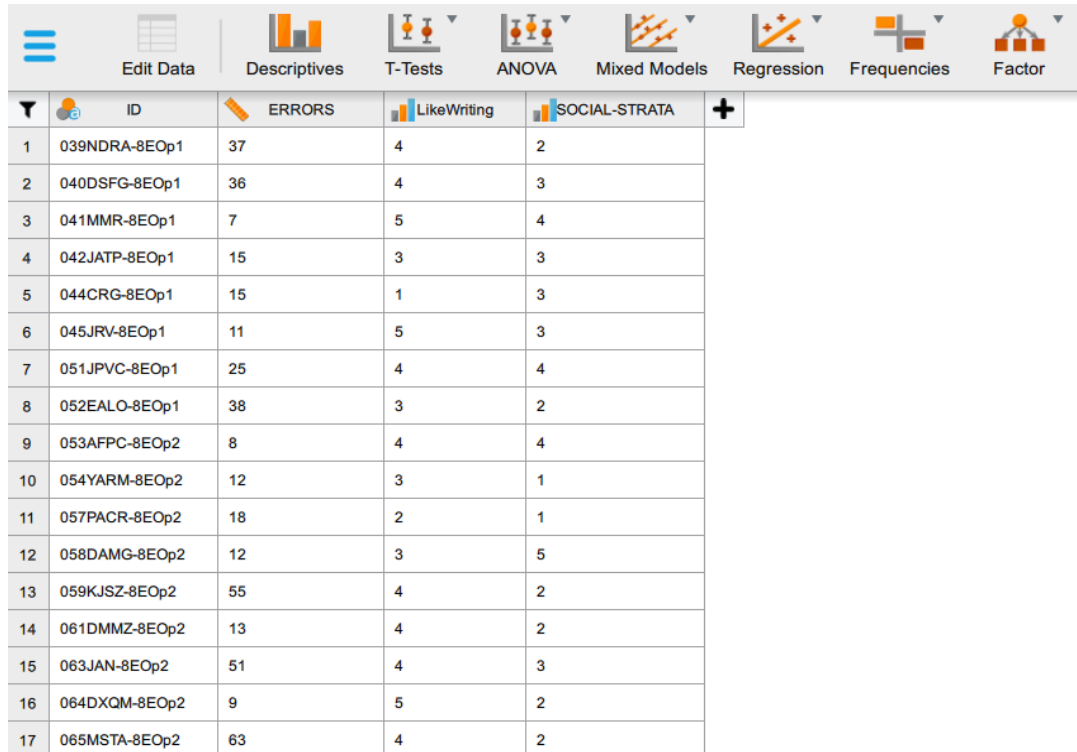
Also, it is relevant to mention that, even though the analyzed corpus consists of 49 texts, only a subgroup of 30 were finally used for the correlational analysis, given that only 30 students responded to the survey.

Figure 9

Data for Correlations in Google Sheets

ID	ERRORS	LikeWriting	SOCIAL-STRATA
039NDRA-8EOp1	37	4	2
040DSFG-8EOp1	36	4	3
041MMR-8EOp1	7	5	4
042JATP-8EOp1	15	3	3
044CRG-8EOp1	15	1	3
045JRV-8EOp1	11	5	3
051JPVC-8EOp1	25	4	4
052EALO-8EOp1	38	3	2
053AFPC-8EOp2	8	4	4
054YARM-8EOp2	12	3	1
057PACR-8EOp2	18	2	1
058DAMG-8EOp2	12	3	5
059KJSZ-8EOp2	55	4	2
061DMMZ-8EOp2	13	4	2
063JAN-8EOp2	51	4	3
064DXQM-8EOp2	9	5	2
065MSTA-8EOp2	63	4	2
066JMS-8EOp2	15	4	3
068CVDR-8EOp2	6	5	3
069ITC-8EOp2	36	3	3
070LLCL-8EOp3	50	4	3
072JPM-8EOp3	31	4	3
073AFSG-8EOp3	26	4	3
074DCEP-8EOp3	27	5	5
076SRP-8EOp3	35	4	4
078DFOP-8EOp3	54	5	2
080LMGS-8EOp3	10	2	3
081MMBB-8EOp3	21	4	3
084JSM-8EOp4	50	4	3
055CCV-8EOp2	13	3	4

After the spreadsheet was completed, the file was saved as comma-separated values (CSV file) and then imported into the software JASP to conduct the statistical and correlational analyzes (Figure 10).

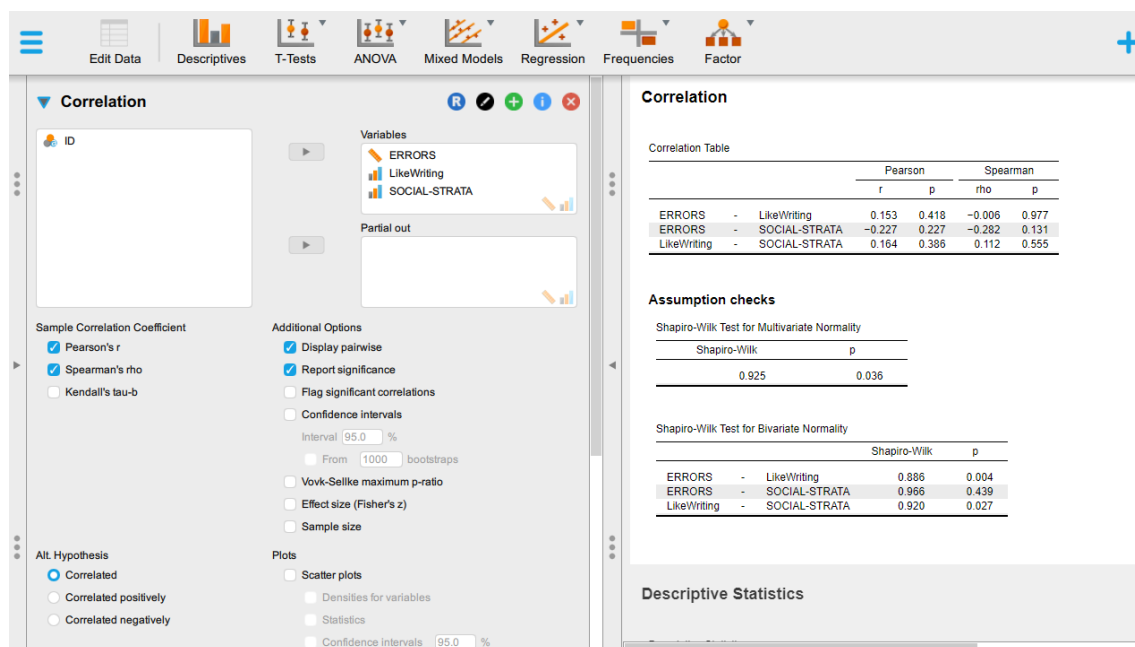
Figure 10*Data Imported into JASP*


	ID	ERRORS	LikeWriting	SOCIAL-STRATA
1	039NDRA-8EOp1	37	4	2
2	040DSFG-8EOp1	36	4	3
3	041MMR-8EOp1	7	5	4
4	042JATP-8EOp1	15	3	3
5	044CRG-8EOp1	15	1	3
6	045JRV-8EOp1	11	5	3
7	051JPVC-8EOp1	25	4	4
8	052EALO-8EOp1	38	3	2
9	053AFPC-8EOp2	8	4	4
10	054YARM-8EOp2	12	3	1
11	057PACR-8EOp2	18	2	1
12	058DAMG-8EOp2	12	3	5
13	059KJSZ-8EOp2	55	4	2
14	061DMMZ-8EOp2	13	4	2
15	063JAN-8EOp2	51	4	3
16	064DXQM-8EOp2	9	5	2
17	065MSTA-8EOp2	63	4	2

Finally, the analysis was conducted. For each pair of variables, both descriptive statistics (including means, modes, medians, standard deviations, and ranges) and bivariate correlations were calculated. The relationships between variables were assessed using a Spearman's rho or a Pearson's correlation coefficient, depending on the results of their corresponding Shapiro-Wilk test for normality (Figure 11). The resulting coefficients and their associated p-values provided a comprehensive overview of the data's characteristics, as presented further in the Results and Discussion section.

Figure 11

Descriptive Statistics and Correlations Carried Out in JASP



3.4. Validity and reliability

This study aligns with the principles of validity and reliability, recognizing their necessary role in ensuring accurate measurement, consistent and dependable results, and the overall credibility and rigor of research contributions to the broader knowledge base. To enhance the validity of our research, we adopted strategies proposed by Shenton (2004). The first strategy involves triangulation, whereby various data collection methods, including error analysis and an online survey, are employed to investigate the same phenomenon, promoting a more comprehensive understanding. Frequent debriefing sessions with professors and other researchers constitute the second strategy, providing a collaborative space for discussions, feedback on proposed actions, and the testing and refinement of the researcher's ideas and

interpretations. These sessions contribute significantly to fostering a comprehensive and unbiased research process. An additional strategy employed in this study is cross-checks between researchers, since we both review transcripts to ensure alignment between the texts and their intended meanings. Moreover, to improve the validation, reliability and accuracy in the analysis, peer review by a professor was incorporated as part of the research strategy as previously described in the methodology (page 44). The expert professor reviewed the full list of interlanguage errors identified across the 49 written compositions, providing critical feedback and confirmation of the categorizations. Through these strategies, we aim to uphold the highest standards of validity and reliability in our research.

3.5. Ethical considerations

In the course of this investigation, we adhered to ethical considerations as outlined in the work of Christensen et al. (2014). These considerations encompassed elements such as obtaining institutional approval, ensuring confidentiality, and preserving anonymity. To ensure confidentiality in this research, the accessibility to the corpus and the students' relevant information collected through the online survey was exclusively restricted to the researchers and the supervising professors overseeing the study. Finally, to preserve the confidentiality of the participants, the researchers made sure that the identities of individuals were not revealed in this study. Additionally, each of the 49 files constituting the analyzed corpus was designated with alphanumeric codes to preclude the use of the names of students who produced the texts.

RESULTS AND DISCUSSION

To begin with the presentation of the results, it is important to remember that the main objective of this study is to find the most common errors in the written interlanguage of eighth-semester English Foreign Language pre-service teachers in the Universidad del Valle, and to explore the correlation between these errors and sociocultural factors, in this case, socioeconomic strata and writing enjoyment.

To guide the results, the study addresses the following specific research questions:

- What are the predominant errors in the written compositions of eight-semester EFL pre-service teachers from the Universidad del Valle?
- To what extent is there a correlation between the participants' social strata and the frequency and types of interlanguage errors observed in their argumentative essays?
- To what extent is there a correlation between participants' level of writing enjoyment and the frequency and types of interlanguage errors present in their argumentative essays?

4.1. Predominant errors in the written compositions

4.1.1. *Results*

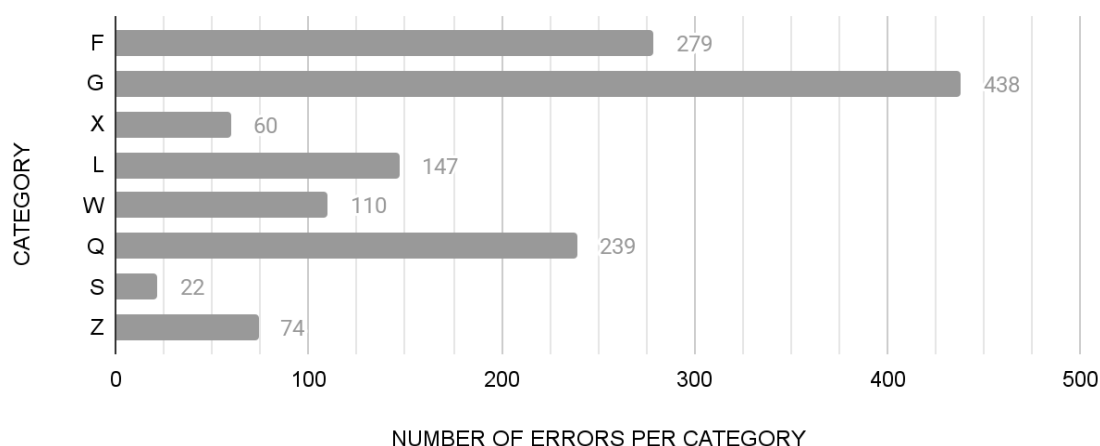
4.1.1.1. Analysis by general categories

The predominant errors identified in the corpus correspond mainly to the categories of grammar (G), form (F), and punctuation (Q). By contrast, other categories such as style (S), infelicities (Z) and lexico-grammar (X) account for a lower quantity of errors in the corpus

(Figure 12). Grammar (G), in general terms, is the category with the highest number of found errors, given that it contains 25 subcategories.

Figure 12

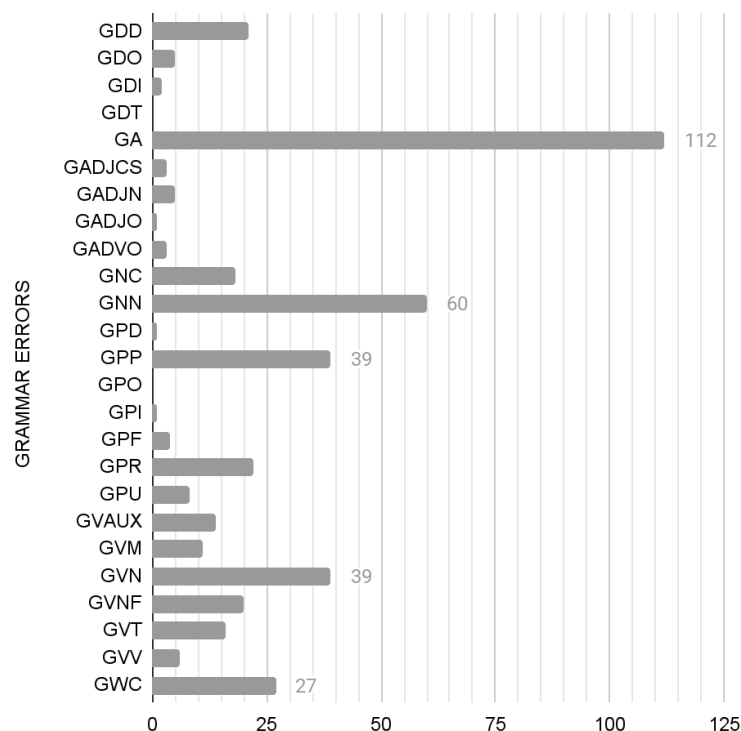
Errors per Category in the Corpus



Within Grammar, errors referring to the use of Articles (GA) are the most frequent among all grammar-related errors, significantly surpassing the other grammar categories with a frequency of 112 appearances (Figure 13). The second most frequent grammatical error type is noun number (GNN), with 60 cases, suggesting that students often confuse singular and plural forms, possibly due to transfer from their first language. Errors related to personal pronouns (GPP) and verb number (GVN) are equally frequent, each occurring 39 times, reflecting challenges in subject–verb agreement and pronoun selection. Finally, word class errors (GWC), with 27 instances, show that students occasionally misuse parts of speech, such as using a noun instead of an adjective.

Figure 13

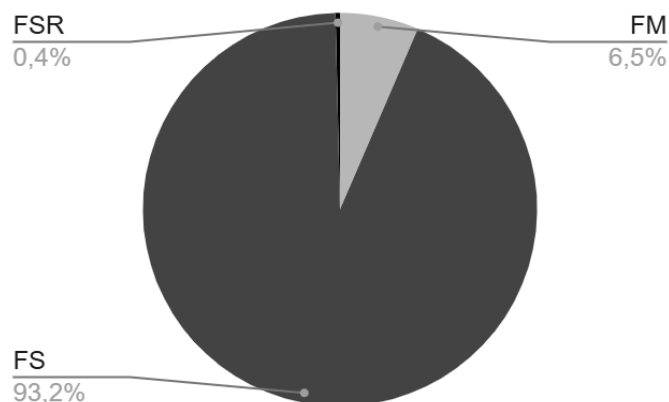
Number of Grammar Errors in the Corpus



To better understand the distribution of error types within the form (F) category, Figure 14 presents the percentage breakdown of each subcategory. Although spelling errors (FS) are the most frequent type of error, the overall form category ranks second in terms of total errors. This suggests that the Form category is characterized almost exclusively by spelling problems, whereas other formal issues are nearly absent from the data.

Figure 14

Percentages of Form Errors in the Corpus

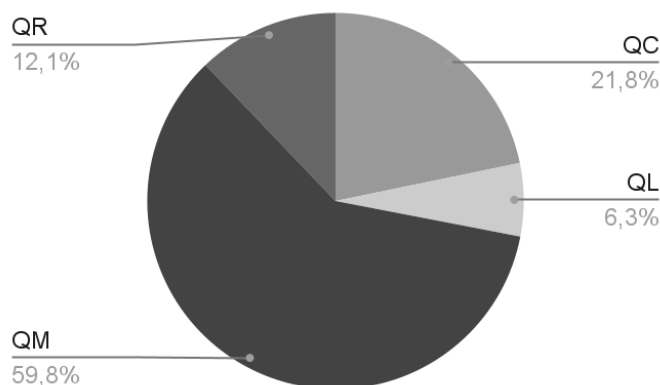


Note. These percentages were calculated in relation to the total number of Form Errors.

On the other hand, although missing punctuation (QM) errors are the second most frequent, the overall punctuation category ranks as third in total errors of the corpus. Figure 15 below shows the distribution of the four types of punctuation errors found. This category is classified into four main subcategories. The most common error is missing punctuation (QM), which represents almost two-thirds of the cases (59.8%). This suggests that students frequently omit necessary commas or other punctuation marks, which affects clarity and sentence structure. QR represents redundant punctuation, for example, the unnecessary repetition of commas. QC indicates confusion of punctuation marks, when one mark is used instead of another (a comma instead of a full stop). Finally, QL occurs when a punctuation mark is mistakenly used in place of a lexical item, usually a conjunction, or when a lexical item is used instead of punctuations. These results reveal that omission is the main challenge in students' writing, while redundancy and confusion occur to a lesser extent.

Figure 15

Percentages of Punctuation Errors in the Corpus.



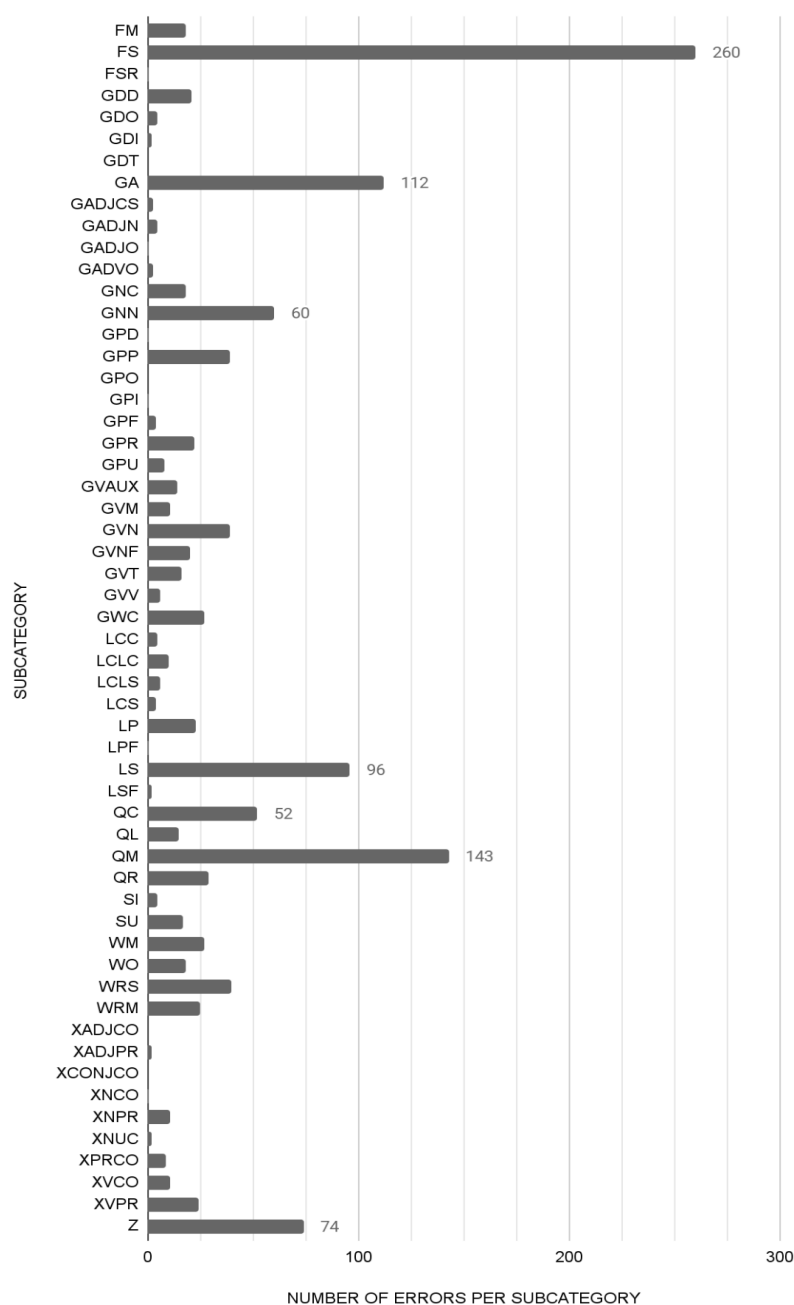
4.1.1.2. Analysis by subcategories

In the following subsections, the main results are examined in greater detail, with a discussion provided for each of the five most important subcategories in order to better understand the sources and implications of the errors found.

As illustrated in Figure 16, the analysis by subcategory provides a more detailed view of the specific errors made by the participants. Among all subcategories, form spelling (FS) shows the highest frequency of errors, indicating persistent challenges with orthography. This is followed by missing punctuation (QM) and grammar articles (GA), both categories representing fundamental elements of English syntax, continuing with errors on single lexical items (LS), related to the vocabulary system, and finally, infelicities (Z), representing slips on pragmatics or appropriacy. These findings indicate that even at advanced stages of learning, fundamental aspects of the language continue to present difficulties for EFL learners. Overall, the distribution of errors in this figure highlights that while some difficulties are indeed concentrated in these core fundamental grammatical and lexical elements, there is a relative scarcity of errors in other categories, implying that students have achieved a strong mastery of them. This suggests that targeted interventions focusing on specific areas are needed.

Figure 16

Errors per Subcategory in the Corpus.



Spelling (FS) errors, errors of a formal nature, were the most common errors. According to the Error Tagging Manual (Dagneaux et al., 2005), an FS error occurs when a word is misspelled, including mistakes in letter arrangement, omitted or incorrect capital letters, and

improper use of hyphens in compound words. It also includes borrowings from other languages, homophones that are incorrectly used, and newly created words that are not derived through standard morphological processes.

The second most common type of error is punctuation. Specifically, the predominant issue was missing punctuation, which is tagged as QM in the manual. Punctuation errors also include redundant punctuation (QR), confusion between punctuation marks (QC), and the incorrect use of punctuation marks in place of lexical items (QL).

Moreover, the third type of predominant errors involves all problems of definite, indefinite or zero article, an error that is categorized as grammatical article (GA) in the manual. This problem is found in the category of grammar, which contains several subcategories besides the one already mentioned. This indicates that learners struggle with the use of definite and indefinite articles, a common difficulty for speakers of languages where articles are used differently or not at all. Some subcategories that are included in this grammar category deal with grammatical aspects such as adjectives, adverbs, nouns, pronouns, verbs and word classes. All these subcategories can be seen in more detail in Table 1.

The fourth subcategory of the most frequently committed errors is related to lexical errors that only occur in single words, and are labelled as LS (lexical single). This subcategory covers, specifically, those lexical errors related to conceptual, collocational and connotative aspects: conceptual errors are committed when an inappropriate word is chosen because its meaning does not match the idea that the learner intended to express; collocational errors involve an incorrect combination of vocabulary that is not typically used together in conventional usage, and connotative errors occur when a word selected by the learner has an associative or emotional

meaning that is unsuitable for the given context. This subcategory also includes errors found in solid and hyphenated compound words.

Finally, infelicities, tagged as Z, is the fifth most common subcategory. According to the Error Tagging Manual (Dagneaux et al., 2005), infelicities, contrary to all the other subcategories, are not real errors, given that the language produced is not grammatically erroneous, but instead considered inappropriate, unnatural or stylistically inadequate. This subcategory includes three types of errors: register problems, questions of political correctness and stylistic problems. First, register problems are produced because a lexical or structural choice does not match the field, mode or tenor of the discourse, for example the use of contractions, informal vocabulary, slang, exclamation marks and archaic forms. Second, problems of political correctness involve word choices that may be considered outdated or exclusionary, such as chairman instead of chairperson or mankind instead of humankind. Third, stylistic problems are committed when a sentence or phrase seems clumsy, foreign-sounding or awkward to the point that a native speaker would want to reformulate it, even if it does not contain any error. Importantly, it must be highlighted that this category should only be used when no other tags are suitable to ensure precision and avoid overgeneralization.

To better illustrate the types of errors found in the analysis, Table 2 presents selected examples of common errors made by learners. Each entry shows the original error, its classification, and the proposed correction according to the Error Tagging Manual (Dagneaux et al., 2005) used in the study.

Table 2

Examples of Predominant Error Types and Their Corrections.

Category	Error	Correction
Form Spelling	<i>nowdays</i>	(FS) <i>nowdays</i> \$nowadays\$
	great wave of <i>internet</i> , where they are	great wave of (FS) <i>internet</i> \$Internet\$, where they are
	information by <i>analizing</i> some disadvantages	information by (FS) <i>analizing</i> \$analyzing\$ some disadvantages
Missing Punctuation	transportation in Cali. First to implement	transportation in Cali. First (QM) 0 \$,\$ to implement
	will cost a lot of money but in a local restaurant	will cost a lot of money (QM) 0 \$,\$ but in a local restaurant
	with this. For example if aiming to build discipline	with this. For example (QM) 0 \$,\$ if aiming to build discipline
Grammar, Articles	But it is necessary to plan <i>an</i> schedule and investigate	But it is necessary to plan (GA) <i>an</i> \$a\$ schedule and investigate
	too much time on Internet using these devices	too much time on (GA) 0 \$the\$ Internet using these devices
	the proficiency level for certain job/situation	the proficiency level for (GA) 0 \$a\$ certain job/situation
Lexical Single	Using these devices has <i>assured</i> that children	Using these devices has (LS) <i>assured</i> \$ensured\$ that children
	buying <i>memories</i> for you and your loved ones	buying (LS) <i>memories</i> \$souvenirs\$ for you and your loved ones
	and <i>specially</i> parents must watch for their kids	and (LS) <i>specially</i> \$especially\$ parents must watch for their kids
Infelicities	as it is logically expected from a regular <i>guy</i> .	as it is logically expected from a regular (Z) <i>guy</i> \$person\$.
	background that we <i>didn't</i> even know that existed	background that we (Z) <i>didn't</i> \$did not\$ even know that existed
	<i>To sum up</i>	(Z) <i>To sum up</i> \$To summarize\$

4.1.2. *Interpretation and discussion of the results*

The first objective of this study is to find the predominant errors in the written compositions of eight-semester EFL pre-service teachers. Through analyzing students' writing production by means of error analysis, it was found that the main errors were: spelling (FS), punctuation (QM) articles (GA), single lexical items (LS), and infelicities (Z). All the previously mentioned errors represent a challenge for learners, since they may affect both the accuracy and clarity of their written communication.

Spelling errors represent the majority of errors in the corpus (19%) with examples such as “health and (FS) well being \$well-being\$” (Text 82) and “the learners, that (FS) wil \$will\$ understand” (Text 47). This type of error has a possible cause, defined by Pardo (2021) as a systematic occurrence due to the lack of contrast between the phonetic system of the mother tongue and the target language. Some other potential causes are, on the one hand, the degree of language distance existing between the spelling systems of English and Spanish, and, on the other, the marked dissonance between the orthographic and phonological systems of English. This latter characteristic results considerably less present in a phonetically consistent language such as Spanish, which constitutes a notable source of errors for Spanish-speaking EFL learners. Also, as many errors of this type were related to the use of vowels, this can be seen in words like “(FS) momentanous \$momentaneous\$” and “(FS) guideance \$guidance\$”, according to Alsahafi (2017), this is due to the inconsistency of English spelling, an issue that is more noticeable in the use of vowels. Moreover, this kind of error might be related to the fact that silent vowels can be problematic for Spanish speakers. Many other cases were due to capitalization mistakes, since students tend to forget some of the rules for capitalizing proper nouns or words after a colon, for

instance this type of error can be seen in sentences like: “problems such as: (FS) Attention \$attention\$” and “need to know, for example: (FS) At \$at\$ midnight”.

The second type of error that stands out is the missing punctuation (QM) with a 10.44%, this can be exemplified in sentences such as “because of that (QM) 0 \$,\$ in many situations” and “on the one hand (QM) 0 \$,\$ it facilitates”. A hypothesis that could explain this issue is that English is a language that uses more periods to divide sentences, while Spanish sentences are often longer and complex, using commas to divide the clauses. This issue has been present in diverse studies such as Alawi (2014), Garcia and Zuluaga (2021) and Chaparro (2022). This last author found in his study that Spanish-speaking students learning German were having difficulties with the use of commas for separating subordinate clauses and direct speech. This issue was due to the ignorance of the rules even in their native language; therefore, they transfer this lack of knowledge to the foreign language they are learning.

The third type of error is related to the use of articles (GA) with an 8.2%. Some examples of this error are: “thanks to (GA) 0 \$the\$ internet” and “because it is (GA) an \$0\$ each day”. It can be seen that students tend to avoid or add definite and indefinite articles erroneously. The possible causes may be due to the interference of their mother tongue, since Spanish is a language that uses articles continuously and more often than English. Some authors who found a high percentage of errors related to articles were Diez (2011) and Pardo (2021). According to this last author, these are “negative transfers” that cause divergences between the first and second languages, causing problems at the sentence and grammatical level (Elejalde and Ferreira, 2017, cited in Pardo, 2021).

The fourth type of error identified in this research is related to the misuse of single words, which are tagged lexical single (LS). Examples of this include: “stressful moment (LS) of \$in\$ the process” and “people in general but (LS) specially \$especially\$ for children”. The majority of errors of this type were related to the wrong use of independent prepositions, as shown in examples such as: “it should be reflected (LS) on \$in\$ the performance” and “most popular activities to do (LS) by \$for\$ many people”. A possible explanation for understanding this issue may be the lack of correspondence that prepositions have in English with their translation in Spanish, since most of them do not have a one-to-one equivalent. Another big amount of LS errors were due to the improper use of words that may have a similar meaning, some examples of that are: “also the typical (LS) plate \$dish\$” and “think about everyone's (LS) expectatives \$expectations\$ and realities”. This problem may be caused by the direct translation of words from one language to another, selecting a word that is not equivalent to what they are trying to say. These results are in line with the results found by Pardo (2021), who found lexical errors in the conceptual use of words and in lexical collocations.

Finally, the last type of errors were infelicities (Z), which are not grammatical errors but inappropriate, unnatural, or stylistically inadequate use of words and expressions. For instance, the following contractions are labeled under this category: “something that (Z) doesn't \$does not\$ happen” and “that is how (Z) you'll \$you will\$ learn more”. Most of the infelicities registered in this study were related to the use of contractions, which are not well seen in formal texts like argumentative essays owing to their informal tone. Another recurrent type of infelicity involved the use of colloquial or non-academic vocabulary, for instance, “access to (Z) tech \$technology\$ devices” and “holiday activity for (Z) kids \$children\$ and adults”. Overall, these findings suggest that learners may still struggle to distinguish between formal and informal

registers in English, highlighting a need for explicit instruction on stylistic appropriateness and register awareness in academic texts.

To conclude this section, it can be said that the main errors identified in the corpus are intrinsically related to students' mother tongue, given that students tend to overgeneralize the knowledge they already possess of their first language, resulting in the production of interlanguage errors since there is a lack of similarities between the phonetic systems and grammar of the languages they use. Also, another cause of these errors can be the lack of knowledge regarding the language rules of both languages, an example of this are the problems related to the use of punctuation signs.

4.2. Correlation between the participants' social strata and the frequency and types of interlanguage errors

4.2.1. Results

In this section, the data obtained from the survey are analyzed, focusing specifically on the students' socioeconomic strata in relation to the frequency and type of errors found in their writing. It is important to note that not all the students responded to the survey; therefore, the data is based on responses from 30 students, as 19 of them did not answer it. To begin with the analysis, the data regarding socioeconomic backgrounds obtained from the survey must be explained for a generalized comprehension.

In Colombia, socioeconomic stratification is organized into six tiers known as *strata*, where stratum 1 represents the lowest income group and stratum 6 represents the highest. As shown in Figure 15, the majority of students who answered the survey belong to the third

stratum, which is generally considered a medium-low socioeconomic level. Table 3 shows that the average socioeconomic stratum of the students is 2.933, displaying that participants' socioeconomic background tends to concentrate between the low and middle class, specifically stratum 3 (lower-middle class) and stratum 2 (low-income class). The statistical results also show a standard deviation of 0.980, evidencing a restrained variability, which indicates scores typically falling between 1.953 and 3.913. With this information, Figure 17 and Table 3 provide a small insight into the students' backgrounds and will allow the exploration of potential relationships between their socioeconomic status and the frequency and types of errors found in their texts.

Figure 17

Students' Socioeconomic Strata.

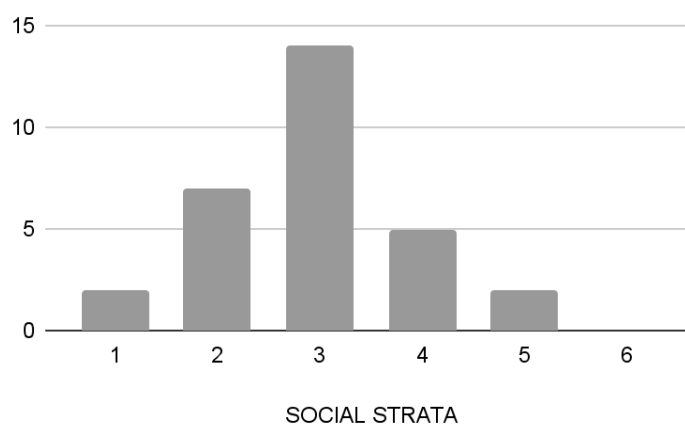


Table 3*Descriptive Statistics for Students' Social Strata*

	Social Strata
Valid	30
Missing	0
Mode	3
Median	3
Mean	2.933
Standard Deviation	0.980
Range	4
Minimum	1
Maximum	5

Once the socioeconomic profile of the participants was established, the next step to continue the analytical process was to determine the appropriate statistical test for establishing correlations. For this, a Shapiro-Wilk test was conducted to assess the normality of the distribution of both the socioeconomic stratum and the total of errors, in order to decide between parametric or non-parametric tests for the subsequent correlation analysis.

The Shapiro-Wilk test is a method to assess if a set of data follows a normal, bell-shaped distribution, and its interpretation depends on the resulting p-value: if a p-value is high (greater than 0.05), it suggests that the data are likely to be consistent with a normal distribution; thus, the null hypothesis of normality is not rejected. Conversely, if a p-value is low (less than or equal to 0.05), then the data are considered to deviate significantly from normality, leading to the rejection of the null hypothesis that the data are normally distributed. According to this clarification, the analysis was carried out.

The assumption of the dataset's normality was tested using the Shapiro-Wilk test. The results, as presented in Table 4, indicated that the data for this variable pair did not deviate significantly from a normal distribution, with a test statistic (W) of 0.966, indicating that the dataset is likely to follow a normal distribution. Additionally, the resulting p-value is 0.439, much higher than the standard significance threshold of 0.05. Based on this finding, it is possible to conclude that there is not enough evidence to reject the null hypothesis of normality of the dataset, justifying the use of a parametric test of correlation. Consequently, Pearson's correlation coefficient was selected for the subsequent analysis.

Table 4

Shapiro-Wilk Test of Normality for Socioeconomic Strata and Total Errors

Variable Pair	Shapiro-Wilk	p-value	Analysis
Socioeconomic Strata - Total Errors	0.966	0.439	Normally distributed

A Pearson's correlation coefficient for datasets with a normal distribution was calculated to examine the relationship between students' socioeconomic strata and the number of writing errors. The results show a very weak negative correlation (Table 5). The test showed a correlation coefficient (r) of -0.227, indicating that there is an inverse relationship, which means that as socioeconomic strata increase, the total number of errors tends to decrease. Nonetheless, the resulting p-value is 0.227, indicating a very weak correlation between these two variables and, thus, manifesting randomness and sample variation as possible causes, rather than a systematic relationship at the population level.

Table 5*Pearson's Correlation Coefficient Between Errors and Socioeconomic Strata*

Variables	Pearson's r	p-value	Analysis
Errors - Socioeconomic strata	-0.227	0.227	Weak negative relationship Not statistically significant

In essence, this relationship was not statistically significant, and no conclusive association was found between the socioeconomic stratum and the students' performance in the writing task. While the data may indicate a very slight tendency for higher-strata students to make fewer errors, this trend is considered too weak and not reliable. The socioeconomic stratum, based on this analysis, is not a predictor of error frequency in this specific context.

4.2.2. Interpretation and discussion of the results

The lack of correlation between socioeconomic status and writing performance in terms of errors might be related to the limited number of survey respondents and the uneven distribution of data, as shown by the Shapiro-Wilk value. For example, only two students reported belonging to the socioeconomic stratum 5, other two to stratum 1, and none of the participants belonged to stratum 6, resulting in a lack of reliability for establishing comparisons and statistical tests. Interestingly, the students from stratum 1 had fewer writing errors than their peers from higher strata, suggesting that socioeconomic strata may not be the only or primary factor influencing writing performance.

In this research, we found that socioeconomic status is not likely to be a significant factor affecting students' performance. These results align with those of Pardo and Quiroz (2021), whose students from lower strata produced fewer mistakes. All of these findings are opposite to

the notion that people from higher strata have a linguistic advantage due to social class privileges that improve language skills, such as access to more opportunities or traveling abroad. This notion can also be opposite to the one proposed by the French sociologist Bourdieu (1979), who states that there is a cultural capital. This concept is related to the access to resources a person can have, such as books, traveling abroad or their parents' education, resources that influence their quality of life and social mobility. Our findings suggest that the supposed linguistic advantage conferred by socioeconomic status is not as prevalent as this concept suggests, and that true linguistic ability may be developed independently of social class.

Continuing with the idea that socioeconomic status may not directly determine language ability (as claimed by Gardner, 2006, and Ellis, 1994), it is important to examine how the socioeconomic background shapes the motivations and views that EFL learners have. The way students perceive the importance of English and its learning can be strongly influenced by their socioeconomic position, which can indirectly affect their language ability. Regarding this specific aspect, Bourdieu's theory of capital does have an important impact and can help with the understanding of this phenomenon. For example, it is possible to consider that English learning tends to be instrumentalized (Usma, 2009), specially by students from lower socioeconomic strata who frequently regard English as a vital resource and an essential cultural capital that they must acquire, given that it can be transformed into economic opportunities and economic capital, thus providing access to social mobility and the improvement of their socioeconomic position (Lee, 2011). In contrast, students from higher socioeconomic backgrounds, who commonly possess inherited cultural and social capital thanks to their family's education, access to resources and opportunities to travel abroad, tend to view English as a natural expectation or a mere additional skill, rather than a necessity or a life-changing tool. This uneven distribution of

language learning motivations across social classes could also be translated into different trends in the intrinsic motivation and the effort that EFL students put into their learning process, depending on their socioeconomic status.

Pardo and Quiroz have the same limitations as us, the scarcity of data, since not all the participants provided their demographic information. In their case, only 515 of the 2,088 students signed the consent form to participate in the research. In their study, Pardo and Quiroz found fewer errors in the first and second strata; errors increased with the following strata and diminished again with the higher level. The same results were found in our study.

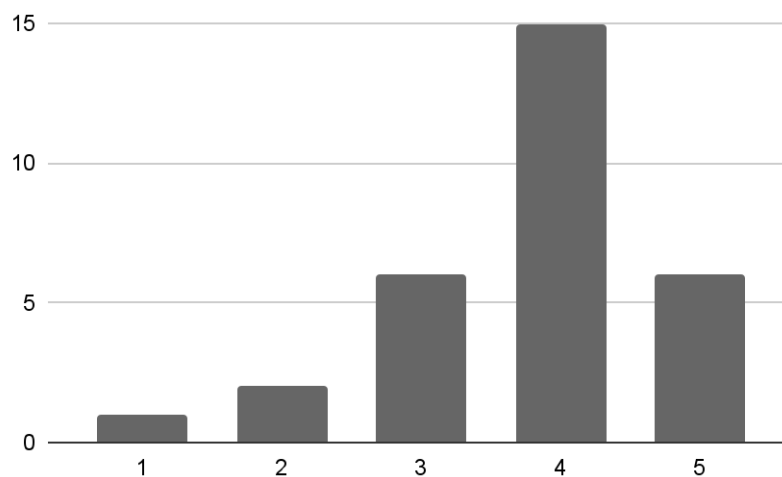
While socioeconomic strata may offer an understanding of the students' writing performance, these findings may not be the only determining factor. Other variables that may influence the level of errors are, for example, the amount of writing that each student produces; this is not taken into account in this study, but it may be another indicator that could change the results, since more production could produce more errors, and less production would significantly change the results.

Students' attitudes toward writing might be another variable playing an important role in determining their academic outcomes. For this reason, the following section explores the relationship between a student's writing enjoyment and the frequency and types of writing errors. The objective is to identify whether a more positive perception of writing correlates with fewer mistakes in their texts.

4.3. Correlation between participants' level of writing enjoyment and the frequency and types of errors

4.3.1. Results

In this section, the objective is to examine whether there is a relationship between students' writing enjoyment and their writing performance, specifically the frequency and types of errors found. Students were asked to rate how much they enjoy writing on a Likert scale from 1 to 5 (1 being the lowest and 5 the highest level of enjoyment). Their responses can be graphically seen in Figure 18, where the majority of students reported relatively high levels of writing enjoyment. Most responses were concentrated around levels 4 and 5, followed by level 3 and a small part of the students selected levels 1 and 2. The mean value of this dataset on the participants' enjoyment of writing is 3.767, suggesting that the attitude towards writing is generally positive, and its corresponding standard deviation is 0.971, representing an important variability (Table 6). These results may have implications for their writing performance and error frequency, thus, this relationship between writing enjoyment and error commission is explored in the next section through a statistical analysis, first by evaluating the normality of the dataset, and then by conducting a correlation analysis.

Figure 18*Students' Writing Enjoyment.***Table 6***Descriptive Statistics for Students' Writing Enjoyment.*

Writing Enjoyment	
Valid	30
Missing	0
Mode	4
Median	4
Mean	3.767
Standard Deviation	0.971
Range	4
Minimum	1
Maximum	5

For evaluating the normality of the dataset, and thus the applicability of statistical methods, a Shapiro-Wilk test was conducted. The results of this test, as shown in Table 7, indicated a Shapiro-Wilk value of 0.886, suggesting that the dataset deviates somewhat from a perfect normal distribution. However, the obtained p-value was 0.004; therefore, the null hypothesis of normality was rejected, and a nonparametric method was employed in the subsequent analysis.

Table 7

Shapiro–Wilk Test of Normality for Writing Enjoyment and Total Errors

Variable Pair	Shapiro-Wilk	p-value	Analysis
Writing Enjoyment - Total Errors	0.886	0.004	Not normally distributed

The Spearman's correlation coefficient for datasets with a non-normal distribution of the variables was used. The results revealed a Spearman's rho of -0.006 with a p-value of 0.977, indicating a negligible and non-significant negative correlation (Table 7). This indicates that students' level of writing enjoyment does not seem to be associated with the number of writing errors they made. In other words, this variable does not influence students' accuracy in this sample.

Table 8

Spearman's Rank Correlation Coefficient Between Errors and Writing Enjoyment

Variables	Spearman's rho (ρ)	p-value	Analysis
Errors - Writing Enjoyment	-0.006	0.977	Negligible relationship Not statistically significant

4.3.2. Interpretation and discussion of the results

The correlation analysis between the variable pair writing enjoyment and error frequency revealed a negligible and statistically non-significant relationship ($\rho = -0.006$, $p = 0.977$), which suggests that, in the case of this sample of EFL pre-service teachers, students' self-reported writing enjoyment does not seem to correlate with their writing accuracy or the number of errors in the argumentative essays they composed. Even if, in the survey, the participants generally reported relatively high levels of writing enjoyment (3.767 on average), this affective factor did not imply any considerable difference displayed in their linguistic performance. Therefore, these findings invite a deeper analysis of the role of learners' affective and motivational variables in EFL writing, as well as on the relevant contextual factors that influence the development of these variables.

From a theoretical perspective, motivational and affective models understand writing enjoyment as a type of intrinsic motivation that fosters engagement and promotes attention to the use of language (Boscolo & Hidi, 2007). However, enjoyment alone might not directly lead to linguistic accuracy, especially in advanced stages of language learning where writing errors often originate because of persistent interlanguage patterns rather than from affective disengagement. This last is supported by theories of writing development (MacArthur & Graham, 2016), which propose that when learners reach higher proficiency levels, their cognitive and linguistic constraints, such as a limited syntactic repertoire and lexical interference, become more decisive on language accuracy than motivational factors. In other words, even when students enjoy writing and have a positive attitude about writing activities, their interlanguage systems might still contain fossilized errors or complex erroneous structures that are hard to overcome.

Bandura's (1997) social cognitive theory also provides a lens to interpret these results. According to Bandura, self-efficacy (rather than enjoyment per se) is the main motivational mechanism influencing language performance. In this case, it is possible that participants' positive affective orientation toward writing did not necessarily reflect a strong sense of writing efficacy. This suggests that enjoyment may coexist with a limited confidence in linguistic accuracy, as found by Setyowati and Sukmawan (2016), who reported that learners with a moderate level of writing motivation still find writing challenging due to grammar and vocabulary. This can be particularly true in foreign language contexts, where learners perceive a dissonance between their ideas and their ability to express them correctly. This could explain the lack of correlation between enjoyment and error frequency previously obtained in the results section, given that, perhaps, students might feel enthusiastic about writing but still struggle with structural and lexical precision due to gaps in their linguistic competences.

Another understanding of these results may be related to the learners' social and instructional environment. As Nolen (2006) emphasizes, writing enjoyment is commonly fostered by meaningful communicative contexts, collaboration, and authentic purposes. The written compositions composing the corpus analyzed in this study, however, were fully academic argumentative essays under test conditions, which may not completely activate the motivational and affective dynamics if compared with personally relevant or creative writing contexts. Therefore, while students reported high enjoyment in general terms, this affective engagement may have been more related to their overall attitude toward writing than to the specific task conditions under which their accuracy was tested.

Finally, these findings suggest two main implications. First, the results highlight the need to distinguish between the learner's general enjoyment of writing and task-specific writing

competence when analyzing performance outcomes. While promoting enjoyment is still an important pedagogical goal, it might not be enough to improve linguistic accuracy on its own. For this reason, it could be necessary to implement explicit instruction and opportunities for reflection focused on language use. Second, these results may suggest to future research the exploration of the interaction between enjoyment, self-efficacy, and strategy use in EFL writing. It is possible that enjoyment indirectly influences performance accuracy through other factors like sustained motivation, careful revision and engagement with feedback, specific factors that were not measured in this study. Overall, the limitations imposed by the *EFL* context (a sociocultural context that is not “friendly” to EFL learning due to, for instance, lack of opportunities to interact in English) might be more influential than individual motivation.

CONCLUSION

To conclude the present study, it is important to highlight the importance of error analysis as an effective tool to understand the challenges that EFL pre-service teachers face. Conducted with the purpose of finding out the correlation between these errors and social strata as well as writing enjoyment, this chapter presents the main findings, the importance for future studies and limitations of the process.

The first research question of this study aimed to identify the predominant errors in written compositions. For the particular case of the Universidad del Valle 8th-semester pre-service teachers, the most frequent errors were spelling mistakes (19%), often related to differences between Spanish and English phonetic and orthographic systems. The consistency of Spanish's phonetic system contrasts with English's irregular spelling, something that especially happens with vowels, leading to possible spelling errors that students may have portrayed in their written compositions. Capitalization mistakes were also common, showing students' confusion about English capitalization rules. The second most frequent error was missing punctuation (10.44%), mainly the omission of commas. This is an error likely coming from Spanish's preference for longer sentences divided by commas, unlike English, a language that uses more periods. Third, article errors (8.2%) were also notable, given that students often omitted or added *a/an/the* incorrectly. This can be attributed to negative transfer from Spanish, since this language uses articles more frequently than English. The fourth most common error identified was the misuse of single lexical items (7.01%). This type of error is mostly related to two issues: the incorrect use of independent prepositions, and wrong application of near-synonyms and semantically related words. The first issue can be caused by the lack of direct one-to-one

equivalents for English and Spanish prepositions, and the second issue, by direct translation, leading to the use of a semantically incorrect word. Finally, the fifth most frequent error were infelicities (5.41%), where language is grammatically correct but used with an unnatural or inadequate style. This error often involves the use of contractions and colloquialisms, indicating that learners struggle with register formality.

Beyond the purely linguistic aspect, this study aimed at correlating written errors and sociocultural factors. The second research question aimed to find the correlation between the errors and socioeconomic strata. The study found no significant correlation, likely due to the small and uneven sample size. Nevertheless, students from lower socioeconomic backgrounds made slightly more mistakes than those from higher socioeconomic backgrounds, suggesting that social class could be a factor influencing language performance, as is often thought. These results align with previous research (Pardo & Quiroz, 2021) and seem to challenge Bourdieu's (1979) concept of cultural capital, which links higher social class to greater linguistic advantages through access to resources like books, trips, and education. Some findings suggest that linguistic ability can be developed independently of socioeconomic status. However, it is important to take into account socioeconomic background, which may still influence learners' context, access to resources, motivation and perception of learning a language. Students from lower strata often see English as a valuable resource for economic improvement and social mobility, while those from higher strata tend to view it as an expected skill or a minor addition.

The third research question aimed to determine the extent of the correlation between the participants' level of writing enjoyment and the frequency of interlanguage errors in their argumentative essays. Firstly, this study found a statistically nonsignificant correlation ($\rho = -0.006$, $p = 0.977$), indicating that, for this sample of EFL pre-service teachers, the relatively

general enjoyment of writing is not directly transformed into greater linguistic accuracy. This may challenge the assumption made in some motivational models that intrinsic enjoyment alone can foster engagement to reduce errors. Conversely, the findings suggest that at advanced learning stages, cognitive and linguistic constraints might be more influential for accuracy than affective factors. These results align with existing theories of writing development, which suggest that a lack of language proficiency can be a bigger obstacle than a lack of motivation (MacArthur & Graham, 2016). Furthermore, the results invite us to distinguish between the general enjoyment of writing and task-specific self-efficacy, as proposed by Bandura (1997), given that students may enjoy writing yet lack confidence in their use of language, a commonly noted situation in foreign language contexts. Also, the academic test-conditioning of the task could have limited the activation of motivational dynamics associated with enjoyment. Consequently, even if promoting enjoyment is an important pedagogical goal, these results imply that it may be insufficient to improve accuracy, acknowledging the need for complementary and explicit language instruction and a research focus on how enjoyment might indirectly influence performance through mediators like self-efficacy, revision strategies, and engagement with feedback.

The findings of this study carry significant implications for both theory and practice in EFL teacher education. Theoretically, the absence of a direct correlation between writing enjoyment and error frequency challenges the simplistic perspective of motivational models regarding linguistic accuracy, suggesting that they must look beyond motivation itself, and start developing more complex, multi-faceted frameworks that integrate aspects such as writer's affective dynamics, cognitive resources and language skills. Furthermore, the mixed findings regarding socioeconomic status complicate a deterministic reading of Bourdieu's cultural capital

theory in EFL writing, which may indicate that linguistic competence can, in certain educational contexts, be cultivated independently of economic background. Now, from a practical perspective, these findings recommend a multidimensional strategy. Pedagogically, instruction should combine writing enjoyment with explicit lessons on common errors like spelling, punctuation, and articles, using a contrastive Spanish-English analysis to take care of errors originating because of linguistic interference from the mother language. Secondly, EFL professors should consider that students from lower-income backgrounds are often driven by English's practical value, which is a powerful motivational tool that can be used in favor of fostering the language learning process. Finally, to better assess motivation and practices related to it, a wider range of writing tasks (for example, personal and creative texts) should be used instead of only depending on essays and other academic compositions in exams.

Regarding limitations, this study found several during its development. The first limitation was the long and labor-intensive process required to analyze all the students' written compositions. Also, the scarcity of answers to the survey and the lack of direct comparability of the findings with those of other studies. The latter is because the research topic (error analysis) is not usually examined alongside variables as those used in this research, given that almost all the studies are of a quantitative nature and do not consider possible sociocultural causes that lead to errors. Also, another point to consider is that the demographics of populations studied in existing literature differed from those in the present study, further constraining the comparability of the obtained findings. An additional methodological limitation related to the error analysis approach lies in the fact that some information related to learners' behaviors was overlooked, for example, avoidance strategies. Another significant limitation to deeply understanding students' writing performance concerns the test itself, since the topic may not have allowed them to fully employ

their knowledge, constraining them from using their whole lexical and syntactic repertoire. Similarly, the exam conditions and the type of text the participants were asked to write could have influenced their perception of the writing activity and their affective dynamics when facing it, which potentially limited their performance as well as the accuracy of the measures and analysis regarding motivation and writing enjoyment.

Now, based on the limitations and findings of this study, several important suggestions for future research emerge. First, to address the sample size limitation, a large-scale, quantitative study with a balanced representation across socioeconomic backgrounds is essential to definitively establish or refute the influence of socioeconomic status in EFL writing. Second, future research should employ a mixed-methods design, integrating error analysis with interviews or stimulated recall to uncover the cognitive and strategic processes behind the errors, and including the gathering of information about the use of avoidance strategies. Also, to further explore the role of affective dynamics, researchers could design studies that compare error rates in test-condition writing versus more authentic and personally meaningful writing tasks, in order to see if enjoyment plays a more significant role in low-stakes environments. Lastly, longitudinal research is needed to track how the relationship between enjoyment, socioeconomic factors, and linguistic accuracy evolves throughout the education program.

In summary, this study provides a valuable and nuanced analysis of the error profiles of EFL pre-service teachers, successfully identifying key linguistic challenges and exploring their relationship with sociocultural and affective factors. Even if the findings indicate that accuracy is not only determined by writing enjoyment or socioeconomic status, this nuanced and deep understanding is a key contribution, because it moves beyond the simplistic explanations to address the intertwined roles of L1 interference, cognitive constraints, socioeconomic

backgrounds and motivation. Consequently, this work offers a starting point for refining some aspects of teacher education and future research in the field.

REFERENCES

- Adebileje, A., Adeleke, A., & Ajilore, T. (2012). Investigating ESL Learners' Socioeconomic Environment on Their Writing Competence in Lagos, Nigeria: Implications for Pedagogy. *Studies in Literature and Language*, 4(2), 122-128.
[10.3968/j.sll.1923156320120402.1020](https://doi.org/10.3968/j.sll.1923156320120402.1020)
- Alawi, T. (2014). An Error Analysis On Academic Writing: A Look Into The Interlanguage Of University Students. *SDSSU Multidisciplinary Research Journal*, 2 (2), 1-5. Retrieved from <https://smrj.nemsu.edu.ph/index.php/SMRJ/article/view/20/18>
- Alsahafi, N. (2017). *An investigation of written errors made by Saudi EFL foundation year students*. [Doctoral dissertation, University of New South Wales]. University of New South Wales Library. <http://dx.doi.org/https://doi.org/10.26190/unsworks/19853>
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice Hall. <https://doi.org/10.4135/9781446221129.n6>
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman and Company.
- Bennett, G. (2010). Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers. *University of Michigan Press ETL*. <https://doi.org/10.3998/mpub.371534>
- Bernstein, B. (2003). *Class, Codes and Control (Volume 1): Theoretical Studies Towards a Sociology of Language* (1st ed.). Routledge. <https://doi.org/10.4324/9780203014035>
- Bourdieu, P. (1986). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood.
<https://doi.org/10.1002/9780470755679.ch15>

- Boscolo, P., & Hidi, S. (2006). The Multiple Meanings of Motivation to Write. In *Writing and Motivation*. Suzanne Hidi and Pietro Boscolo (Eds.), (pp. 1-14). Brill.
https://doi.org/10.1163/9781849508216_002
- Chaves, O. (2011). *Errors in written argumentative essays of intermediate EFL students at the Universidad del Valle*. [Master's report, Universidad del Valle]. Biblioteca Digital - Universidad del Valle. <https://hdl.handle.net/10893/21593>
- Chaparro, L. (2022). *Análisis de errores en la interlengua de aprendices colombianos de alemán el papel de la transferencia lingüística*. [Doctoral dissertation, Universidad pedagógica y Tecnológica de Colombia, Facultad de Ciencias de la Educación].
<http://repositorio.uptc.edu.co/handle/001/8968>
- Christensen, L., Johnson, B., & Turner, L. (2014). *Research methods, design, and analysis*. Pearson.
- Christensen, L., Johnson, B., & Turner, L. (2011). *Educational research*. Pearson.
- Corder, S. (1981). Error Analysis and Interlanguage. *Applied Linguistics*, 2. Oxford University Press
- Council of Europe. (2017). *Global Scale - Table 1 (CEFR 3.3): Common Reference levels*. Common European Framework of Reference for Languages (CEFR). Retrieved July 8, 2025, from <https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale>
- Creswell, J. (2012). *Educational research*. Pearson.
- Dagneaux, E., Denness, S., Granger, S., & Meunier, F. (1996). *Error tagging manual version 1.1*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain.

- De Bernardi, B., & Antolini, E. (2006). Chapter 10: Fostering Students' Willingness and Interest in Argumentative Writing: An Intervention Study. In *Writing and Motivation*. Brill.
https://doi.org/10.1163/9781849508216_011
- Departamento Administrativo Nacional de Estadística [DANE]. (2021). *Estratificación socioeconómica para servicios públicos domiciliarios*. Government of Colombia.
 Retrieved July 9, 2025, from <https://www.dane.gov.co/en/citizen-service/information-services/socioeconomic-stratification#preguntas-frecuentes>
- Díez-Bedmar, M. (2011). Spanish pre-university students' use of English: CEA results from the University Entrance Examination. *International Journal of English Studies*, 11(2), 141–158. <https://doi.org/10.6018/ijes/2011/2/149681>
- Dölek, O., & Hamzadayi, E. (2018). Comparison of Writing Skills of Students of Different Socioeconomic Status. *International Journal of Progressive Education*, 14(6), 117-131.
<https://eric.ed.gov/?id=EJ1201802>
- Donoso, E. & Gómez, R. (2018). Errores en discurso escrito en inglés por futuros profesores de inglés como lengua extranjera. *Literatura y lingüística* (38), 175-193.
<https://dx.doi.org/10.29344/0717621x.38.1632>
- Ellis, R. (1994). *The study of second language acquisition*. Oxford University Press, 10(1), 53-55. <https://doi.org/10.5785/10-1-248>
- Gardner, R. (2006). The socio-educational model of second language acquisition. *EUROSLA Yearbook*, 6(1), 237–260. <https://doi.org/10.1075/eurosla.6.14gar>
- Garrido, G., and Romero, C. (2012). Errors in the use of English tenses. *Íkala, Revista de Lenguaje y Cultura*, 17(3), 285–296. <https://doi.org/10.17533/udea.ikala.8724>

- García, J. & Zuluaga, J. (2021). Análisis de la interlengua escrita de estudiantes de nivel intermedio de español como lengua extranjera. *Revista Lengua y Habla*, 25, 344-378.
- Granger, S. (2003). The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research. *TESOL Quarterly*, 37(3), 538–546. <https://doi.org/10.2307/3588404>
- Granger, S. (2008). *Learner Corpora*. Wiley Online Library. <https://doi.org/10.1002/9781405198431.wbeal0669>
- Hoyos, L. & Roldán, J. (2016). *Análisis de errores sintácticos en inglés por interferencia del español en estudiantes de la Licenciatura en Lenguas Extranjeras de la Universidad del Valle*. [Trabajo de grado, Universidad del Valle]. Biblioteca digital - Universidad del Valle. <https://bibliotecadigital.univalle.edu.co/entities/publication/8b242fb3-aa46-441c-9c13-2e224173542d>
- Hutchinson, J. (1996). *UCL Error Editor*. Louvain-la-Neuve: Centre for English Corpus Linguistics, Université Catholique de Louvain.
- JASP Team (2023). *JASP* (Version 0.18.1) [Computer software].
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. Routledge. <https://doi.org/10.4324/9781315843674>
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(2), 205–254. <https://doi.org/10.1017/S0954394500000338>
- Lee, C.G. (2011). English Language and Economic Growth: Cross-Country Empirical Evidence. *Journal of Economic and Social Studies*, 1, 5-20. <https://omeka.ibu.edu.ba/items/show/2627>

- Lemos, C. (2022). *Error Analysis in Written Productions of EFL Elementary Students at a Bilingual Private School in Cali*. [Master's report, Universidad Icesi]. Biblioteca Digital - Universidad Icesi. https://repository.icesi.edu.co/biblioteca_digital/bitstream/10906/95226/1/T02467.pdf
- Lenneberg, E. (1967). The Biological Foundations of Language. *Hospital Practice*, 2(12), 59–67. <https://doi.org/10.1080/21548331.1967.11707799>
- Lozano, C. & Mendikoetxea, A. (2013). Learner corpora and second language acquisition: the design and collection of CEDEL2. In A. Díaz-Negrillo, N. Ballier & P. Thompson (Ed.), *Automatic Treatment and Analysis of Learner Corpus Data* (pp.65-100). John Benjamins Publishing Company. <https://doi.org/10.1075/scl.59.06loz>
- MacArthur, C. & Graham, S. (2016). Writing Research from a Cognitive Perspective. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*, 2, pp. 24–40. The Guilford Press.
- Mariappan, R., Tan, K., Yang, J., Chen, J. & Chang, P. (2022). Synthesizing the Attributes of Computer-Based Error Analysis for ESL and EFL Learning: A Scoping Review. *Sustainability*, 14(23), 15649. <https://doi.org/10.3390/su142315649>
- McEnery, T. & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511981395>
- Myles, F. (2005). Interlanguage corpora and second language acquisition research. *Second Language Research*, 21(4), 373–391. <https://doi.org/10.1191/0267658305SR252OA>
- Nemser, W. (1971). Approximative systems of foreign language learners. *International Review of Applied Linguistics in Language Teaching*, 9(2), 115. <https://doi.org/10.1515/iral.1971.9.2.115>

- Nolen, S. (2006). The Role of Literate Communities in the Development of Children's Interest in Writing. In *Writing and Motivation*. Suzanne Hidi and Pietro Boscolo (Eds.), (pp. 241–256). Brill.
- Pajares, F., Valiante, G., & Cheong, Y. (2006). Chapter 8: Writing Self-Efficacy and Its Relation to Gender, Writing Motivation and Writing Competence: A Developmental Perspective. In *Writing and Motivation*. Brill. https://doi.org/10.1163/9781849508216_009
- Pardo, M. (2021). El análisis de errores como evidencia de la interlengua de los estudiantes de inglés como lengua extranjera. *Literatura y Lingüística*, (43), 281-307.
<https://dx.doi.org/10.29344/0717621x.43.2836>
- Pardo, M. & Quiroz, G. (2021). Género y estrato socioeconómico, ¿factores determinantes en el número y tipo de errores escritos del inglés como lengua extranjera? *Lenguaje*, 49(2), 394-437. <https://doi.org/10.25100/lenguaje.v49i2.10540>
- Raoofi, S. & Maroofi, Y. (2017). Relationships among motivation (self-efficacy and task value), strategy use and performance in L2 writing. *Southern African Linguistics and Applied Language Studies*, 35(3), 299-310, [10.2989/16073614.2017.1391706](https://doi.org/10.2989/16073614.2017.1391706)
- Sabti, A., Md Rashid, S., Nimehchisalem, V., & Darmi, R. (2019). The Impact of Writing Anxiety, Writing Achievement Motivation, and Writing Self-Efficacy on Writing Performance: A Correlational Study of Iraqi Tertiary EFL Learners. *SAGE open*, 9(4), 1-13. <https://doi.org/10.1177/2158244019894289>
- Scott, M. (2008). *WordSmith Tools* (Version 5.0.0.334) [Computer software]. Lexical Analysis Software Ltd. <https://www.lexically.net/wordsmith/>
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching*, 10(1-4), 209. <https://doi.org/10.1515/iral.1972.10.1-4.209>

- Setyowati, L. & Sukmawan, S. (2016). EFL Indonesian Students' Attitude toward Writing in English. *Arab World English Journal (AWEJ)*, 7(4), 365- 378.
<https://dx.doi.org/10.2139/ssrn.2898636>
- Shenton, A. (2004). Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*, 22(2), 63-75. <https://doi.org/10.3233/EFI-2004-22201>
- Usma, J. (2009). Education and Language Policy in Colombia: Exploring Processes of Inclusion, Exclusion, and Stratification in Times of Global Reform. *Profile: Issues in Teachers' Professional Development*, 11(1), 123–141. <https://revistas.unal.edu.co/index.php/profile/article/view/10551>
- Villayandre, M. (2008). Lingüística con corpus (I). *Estudios Humanísticos. Filología* (30), 329-349. <https://doi.org/10.18002/ehf.v0i30.2847>
- Yakut, I. (2022). The Basics of Corpus Linguistics: An Introduction for Beginners. In H. Yilmaz, E. Yucel, & M. S. Ozturk (Eds.), *A General Perspective on Foreign Language Teaching VI* (pp. 176-190). Çizgi Kitabevi. Retrieved from: <https://www.cizgikitabevi.com/kitap/1564-yabanci-dil-ogretimine-genel-bir-bakis-vi>
- Zare-ee, A. (2010). Sustained effects of socioeconomic status on proficiency and academic writing in English as a foreign language. *Teaching English Language*, 4(2), 117-140.
[10.22132/tel.2010.66109](https://doi.org/10.22132/tel.2010.66109)

APPENDIX

Encuesta de variables individuales y socioculturales que podrían incidir en los errores en la escritura en inglés como lengua extranjera

Gracias por participar en la investigación de interlengua de la Escuela de Ciencias del Lenguaje de la Universidad del Valle. Sus textos del examen de competencias (Semestre IV u VIII) ya han sido analizados y ahora le pedimos que conteste unas preguntas que se refieren a algunas variables individuales (por ej., gusto por la escritura en inglés) y socioculturales (por ej., estrato social) que pueden estar relacionadas con el desarrollo de la escritura en inglés. Le recordamos que la información aquí consignada sólo se utilizará para propósitos académicos (varios trabajos de grado de la Licenciatura en Lenguas extranjeras de la Universidad del Valle) y se protegerá la identidad individual de los participantes. Para mayor información sobre este proyecto de investigación pueden contactar al profesor Orlando Chaves, orlando.chavez@correounivalle.edu.co y/o a la profesora Claudia Toro al correo claudia.toro@correounivalle.edu.co.

1. Correo * _____

2. Nombre * _____

Factores Individuales

3. Edad *

Marca solo un óvalo por fila.

15-20 ☐

21-25 ☐

26-30 ☐

Mayor de 30 ☐

4. Sexo*

Marca solo un óvalo por fila.

Femenino ☐

Masculino ☐

5. ¿Cuál es su lengua materna?*

Marca solo un óvalo por fila.

Español ☐

Inglés (Creole San Andrés) ☐

Lengua de señas ☐

Lengua Indígena Colombiana ☐

Otro: _____

Si escogió la opción “Lengua Indígena” u “Otra” por favor indique cuál.

4. Califique su gusto por los siguientes aspectos de la lengua inglesa *

(1 -muy poco- a 5 -muchísimo-)

	1 (Muy poco)	2 (Poco)	3 (Neutral)	4 (Mucho)	5 (Muchísimo)
El idioma Inglés, en general	☐	☐	☐	☐	☐
La escritura en Inglés	☐	☐	☐	☐	☐
La lectura en Inglés	☐	☐	☐	☐	☐
La gramática del Inglés	☐	☐	☐	☐	☐
La cultura de los países angloparlantes	☐	☐	☐	☐	☐

5. ¿Cuáles otras lenguas habla o estudia?*

Marca solo un óvalo por fila.

Lengua	
Francés	☐
Portugués	☐
Alemán	☐
Italiano	☐
Japonés	☐
Mandarín (Chino)	☐
Coreano	☐
Ruso	☐
Otra	☐
Ninguna	☐

6. ¿Cuál es la lengua que predomina en sus dispositivos electrónicos de uso regular (celular, computador, tablet, kindle, etc.)?*

	Español	Inglés	Francés	Portugués	Otra lengua
Celular	☐	☐	☐	☐	☐
Computador	☐	☐	☐	☐	☐
Tablet	☐	☐	☐	☐	☐
Kindle	☐	☐	☐	☐	☐
Otro	☐	☐	☐	☐	☐

7. ¿En qué países de habla inglesa ha vivido? (Puede señalar varios, de ser el caso)*

Marca solo un óvalo por fila.

	País de habla inglesa
Ninguno	☐
The USA	☐
Canadá	☐
UK	☐
Australia	☐
New Zealand	☐
Otro	☐

8. ¿En qué países de habla inglesa ha estudiado? (Puede señalar varios, de ser el caso)*

Marca solo un óvalo por fila.

	País de habla inglesa
Ninguno	☐
The USA	☐
Canadá	☐
UK	☐
Australia	☐
New Zealand	☐

Otro

☐

Factores socioculturales

9. Estrato Socioeconómico*

Marca solo un óvalo.

☐

Estrato 1

(bajo-bajo)

☐

Estrato 2(bajo)

☐

Estrato 3 (medio-bajo)

☐

Estrato 4 (medio)

☐

Estrato 5 (medio-alto)

☐

Opción 6 (alto)

10. ¿Estudió en un colegio bilingüe? *

Marca solo un óvalo.

Sí

☐

No

☐

11. ¿Estudió o estudia en un instituto de lenguas? *

Marca solo un óvalo.

Sí

☐

No

☐

12. ¿En qué contextos fuera del salón de clase se comunica usualmente en inglés?*

Marca solo un óvalo por fila.

Contexto	
Redes Sociales	☐
Amigos	☐
Familia	☐
Profesores	☐
Trabajo	☐
Otro	☐