**6th International Conference on Humanistic Studies**

**Workshop 4 Language Teaching and Development of Competences in the Educational Field**

# **Approaching the Assessment of Written Production through Coh-Metric Principal Component Indices**

***Title***

***Acercamiento a la evaluación de la producción escrita a través de los índices de componentes principales de Coh-Metrix***.

**Name and Surname1, Name and Surname2, …**

1- Humberto Minoso Machado. Universidad Central Marta Abreu de Las Villas, Cuba. E-mail: humbertomm@uclv.edu.cu

2- Tania B. Machado Armas. Universidad Central Marta Abreu de Las Villas, Cuba.

E-mail: taniama@uclv.edu.cu

3-Juana Idania Pérez Morales. Universidad Central Marta Abreu de Las Villas, Cuba.

E-mail: juanap@uclv.edu.cu

**Abstract:**

The learning of English in the UCLV context has encouraged teachers and researchers to find out different ways to enhance the development of proficiency in written production. Assessment is one of the areas in which washback has been expected especially in written production. The present study intends to show new assessment strategies in written production that could enhance teaching and learning success in current courses. Thus the aim of this study is to provide teachers with more evidence to understand better the assessment of written production in English, especially for first-year English language students at Universidad Central Marta Abreu de Las Villas. To fulfill the aim proposed, three Coh-Metrix principal components were analyzed. Through the analysis of descriptive statistics, the authors of this propose the necessary inferences to assess the quality essay writing. The authors followed an action research approach to collect and process quantitative data using Coh-Metrix 3.0 to arrive to proper conclusion based on the indices obtained from two datasets. The results showed the effectiveness of the scores for all indices of three principal components of the multilevel theoretical framework of text and discourse comprehension as a tool to measure the quality of the essays written by the students. From a practical perspective, this study provides evidence for teachers to assume a quantitative way to assess written production by considering three essential Coh-Metrix indices, which have proved to make distinctions between low and high quality essays

***Resumen:*** *El aprendizaje del inglés en el contexto de la Universidad Central Marta Abreu de Las Villas ha incentivado a docentes e investigadores a buscar diferentes formas de potenciar el desarrollo de competencias en la producción escrita. La evaluación es una de las áreas en las que se espera una retroalimentación positiva, especialmente en la producción escrita. El presente estudio pretende mostrar nuevas estrategias de evaluación en la producción escrita para potenciar el éxito en la enseñanza y el aprendizaje. El objetivo de este estudio es proporcionar a los docentes más evidencias para comprender mejor la evaluación de la producción escrita en inglés; en especial, para los estudiantes de primer año de lengua inglesa de la UCLV. Para lograr el objetivo propuesto, se han analizado tres componentes principales de Coh-Metrix. A través del análisis de estadísticas descriptivas, los autores de este trabajo proponen las inferencias necesarias para evaluar la calidad de la escritura. Los autores siguieron un enfoque de investigación de acción para recopilar y procesar datos cuantitativos utilizando Coh-Metrix 3.0 y llegar a una conclusión adecuada basada en dos conjuntos de datos. Los resultados mostraron la eficacia de todos los índices de los tres componentes principales del marco teórico multinivel de comprensión textual y discursiva usados para medir la calidad de la producción escrita. Desde una perspectiva práctica, este estudio proporciona evidencias para que los profesores asuman de forma cuantitativa la evaluación de la producción escrita considerando tres índices Coh-Metrix que demuestran diferencias entre escritores de baja y alta calidad.*

**Keyswords:**

Syntactic complexity; Lexical Diversity; Referential cohesion, Integrated; Independent

***Palabras Claves:*** *Complejidad sintáctica; diversidad lexical; Cohesión referencial*; Integrado; Independiente

# 1. Introduction

Writing skill is crucial nowadays. Becoming a proficient writer is one of the major objectives of many students, especially for those who want to become linguists, teachers or interpreters. The Common European Framework of Reference for Languages (CEFR), descriptors for assessing writing skills has been of considerable interest in the context of the major English Language Studies. However, in the context of Universidad Central Marta Abreu de Las Villas, relatively little research has been done in the field of The CEFR descriptors for assessing writing skills, compared with research in other areas of language teaching and learning.

Further research on the assessment of the written production is therefore necessary to understand better, how L2 learners improve this essential production component of language acquisition and learning. Considering the assessment criteria to assess written production, significant explanations can be provided if additional measuring tools are implemented from a multilevel analysis of text characteristics. Coh-Metrix seems to be the ideal tool to help teachers understand better the assessment of written production.

This study was conducted to provide teachers with more evidence to understand how to assess written production during the learning of writing in English for first-year English language students at Universidad Central Marta Abreu de Las Villas. Three Coh-Metrix principal components were analyzed which provided the descriptive statistics necessary to improve the quality of the assessment criteria for written production learning.

To achieve the goal to understand the extent of helpfulness when Coh-Metrix principal components are analyzed, the following research question is raised.

* To what extent do three Coh-Metrix principal components help teachers understand better the assessment of written production?

The purpose of this paper is to analyze a sample of independent essays and a sample of integrated ones to identify the quality of essays using the multilevel framework of text and discourse comprehension.

### 1.1 The Common European Framework of Reference Assessment Criteria

The top CEFR descriptor for assessing coherence and cohesion **refers to** a well-organized and coherent textwhen the candidate uses a variety of cohesive devices and organizational patterns to cause a good effect on target readers. The following descriptor also refers to using linking words and cohesive devices. For example, *Text* ***generally*** *organized and coherent, using a variety of linking words and cohesive devices. Text is connected and coherent, using* ***basic*** *linking words and a* ***limited number*** *of cohesive devices.*

The CEFR Companion Volume (2018) defines coherence and cohesion as the way in which the separate elements of a text are interwoven into a coherent whole by exploiting linguistic devices such as referencing, substitution, ellipsis and other forms of textual cohesion, plus logical and temporal connectors and other forms of discourse markers (“Common European Framework of Reference for Languages: Learning, teaching, assessment: Companion Volume with New Descriptors”, 2018). Although this document mentions linguistic devices such as referencing, substitution, ellipsis and other forms of textual cohesion, they are only expressed as cohesive devices in the Cando statements for B2 level. However, the descriptors for assessing coherence and cohesion do not specify such linguistic devices. Thus, Coh-Metrix can help novel teachers identify problems and propose solutions at a finer grain size within a multilevel framework of text and discourse comprehension.

### 1.2 Multilevel framework of text and discourse comprehension

Grasser and McNamara (2011) proposed a theoretical multilevel framework consisting of five fundamental levels of language and discourse: *words*, *syntax*, the explicit *textbase*, the *situation model*, and the discourse *genre and rhetorical structure.* This study includes the first three levels only.

#### **1.2.1 Words**

The analysis of words on multiple characteristics that have relevance in meaning construction is paramount to assess the students’ written production effectively. However, the characteristic features that make distinction in terms of writing quality are word frequency as measured by CELEX word frequency (logarithm including all words), and lexical diversity (Measure of Textual Lexical Diversity).

#### **1.2.2 Syntax**

Grasser et al. (2006) placed great emphasis on how complex syntactic structures tax the reader’s working memory. They also refer to the importance of considering embedded words before a main verb as the index of syntactic complexity that best distinguish low and high quality essay (Graesser, Cai , Louwerse, & Daniel , 2006).

#### **1.2.3 The explicit textbase**

The explicit textbase, as defined by Kintsch (1998), consists of the explicit ideas in the text: the meaning rather than the surface code of wording and syntax (Kintsch, 1998). Another important constituent method of connecting propositions, clauses, and sentences in the explicit textbase is co-reference. Referential cohesion occurs when a noun, pronoun, or noun phrase refers to another constituent in the text (Grasser & MacNamara, 2011). They also estimate that, this characteristic is the extent to which explicit content words and ideas in the text are connected with each other as the text unfolds. The noun phrases are important for providing co-reference and bridging the explicit clauses and sentences in the explicit textbase. A referential cohesion gap occurs when a sentence has few if any words that overlap with previous sentences.

# **2. Methodology**

This study analyzed 40 independent essays and 32 integrated ones. The participants involved in this study represented a group of 22 students taking a distance course of written production during the Covid-19 pandemic.

The essays were processed using Coh-Metrix 3.0. The data collected was analyzed based on three theoretical levels (Words, syntax and the explicit textbase) which had proved to make distinctions between low- and high-proficiency essays in previous studies. The indices for the three chosen levels are respectively the following: CELEX Log frequency for all words, syntactic complexity expressed in left embeddedness, words before main verb, mean, and lexical diversity MTLD all words,

We also considered other indices that gave distinctive values among the essays processed in both datasets (independent and integrated essays): Referential cohesion z- score, syntactic simplicity z-score, and content word overlap value.

# **3. Results and Discussion**

## 3**.1 Results**

### 3.1.1 Words

The CELEX logarithm frequency for all words, mean, generally keeps stable from 3.25 to a lower frequency around 2.71 in the case of independent essays. The higher the mean log frequency values the easier to decode the text by the part of the reader. Thus, lower values can be predictive of writers using words less frequently used in the language. However, the integrated essays showed descriptive statistics with most of the log frequency for all words, mean value below 3.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Datasets | Mean | Standard deviation | Minimum | Maximum |
| Independent essays | 2.99 | 0.139 | 2.71 | 3.25 |
| Integrated essays | 2.93 | 0.104 | 2.71 | 3.13 |

Table 1 CELEX logarithm frequency for all words, mean (own elaboration)

Essay 9 showed the lowest CELEX log frequency mean (2.71) of all the independent essays processed. In the case of integrated essays, essay 26 yielded the same lowest mean of all the dataset.

Graph 1 CELEX Log frequency for all words, mean (own elaboration)

### 3.1.3 The explicit textbase

The study showed that referential cohesion z-score could provide teachers with evidence to identify weakness and strength on coherence and cohesion descriptor when compared to only all connectives incidence. The following comparison of the results for the independent and integrated essays’ datasets demonstrated the importance of permanent surveillance of students’ writing outputs to see how the descriptor *coherence and cohesion* behaves in the learning process.

3.3.1 Referential cohesion

Values of the Referential cohesion z-score below 0 indicate acceptable expected results whereas indices above zero require particular analysis. Readers can notice in graph 4 both datasets showed dispersion for referential cohesion.

|  |  |  |
| --- | --- | --- |
| Essay type | Referential cohesion z-score Mean  |  Standard deviation |
| Independent essays | -0,328 | 0,614 |
| Integrated essays  | -0,171 | 0,336 |

Table 3 Referential cohesion **z-core** (own elaboration)

Graph 4 Referential cohesion z-score (own elaboration)

### 3.1.2 Syntactic complexity

McNamara and et al. (2010) considered that the number of words before the main verb was the Coh-Metrix index that distinguished best from essay quality among other indices of syntactic complexity. Graph 2 illustrates the indices in two datasets with few fluctuations above the 6 mean values. Descriptive statistics for syntactic complexity are shown in table 2

|  |  |  |
| --- | --- | --- |
| Descriptive statistics  | SYNLE for dataset 1 *Independent essays* | SYNLE for dataset 2 *Integrated essays* |
| Mean | 5.26 | 5.28 |
| Standard deviation | 1.46 | 1.36 |
| Maximum | 10.04 | 10.04 |
| Minimum | 2.58 | 3.15 |

Table 2 Left embeddedness, words before main verb, mean SYNLE (own elaboration)

Graph 2 Syntactic complexity Left embeddedness, words before main verb, mean (own elaboration)

Nevertheless, they claim that z-score index for syntactic simplicity is also an indicator of essay quality. In the graph below, readers can easily recognize that values below 0 indicate the expected values rather than those essays with values greater than 0, which requires particular assessment (e.g. text 10).

Graph 3 Syntactic simplicity z-score (own elaboration)

### 3.1.4 Lexical diversity

McNamara and McCarthy (2010) affirmed that greater lexical diversity of words in a text means that wider varieties of words are used across the text, which is associated with more challenging text. Lower frequency words in a text means that the words are less familiar and less accessible to the reader (McNamara, Mccarthy, & Crossley, 2010).

The lexical diversity descriptive statistics for the two datasets are shown in the following table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Datasets | Mean | Maximum | Minimum | Standard deviation |
| Independent essays | 92,43 | 126.81 | 51.75 | 18,91 |
| Integrated essays | 91,43 | 130.79 | 59.69 | 17,96 |

Table 4 Lexical diversity (Descriptive statistics (own elaboration)

Graph 5 Measure of Textual Lexical Diversity (MTLD) (own elaboration)

### 3.1.5 Content word overlap

Guoa et al. (2013) found that low content word overlap scores indicate a more proficient essay writer in the case of independent essay. However, integrated essay writing show a positive correlation between high content word overlap and high proficient writers (Guoa, Crossley, & McNamara, 2013).In this study, the descriptive statistics for content word overlap corroborates their point of view as can be seen in the following table:

|  |  |  |
| --- | --- | --- |
|  | Independent essays | Integrated essays |
| Mean | 0.083 | 0.123 |
| Standard deviation | 0.025 | 0.033 |
| Maximum | 0.137 | 0.177 |
| Minimum  | 0.041 | 0.039 |

Table 5 Content word overlap scores (own elaboration)

 The graph below illustrates, both, the independent and the integrated datasets scores for content word overlap

Graph 6 Content word overlap, all sentences, proportional mean(own elaboration)

# **3.2 Discussion**

This study has investigated whether Coh-Metrix three principal components were predictive measures of essay quality or not. The literature review indicated that three principal components were predictive measures of essay quality. These measures made distinctions between low and high proficient writers in terms of word frequency, syntactic complexity and lexical diversity. The indices chosen for word frequency as a result of McNamara, Mccarthy and Crossley’s research were respectively CELEX logarithm for all words, number of words before the main verb, and the Measure of Textual Lexical Diversity (McNamara, Mccarthy, & Crossley, 2010). This study also demonstrated that those essays that deviated from the mean value of the dataset required detailed assessment, especially those that did not demonstrate the expected indices.

## **3.2.1 Words**

Words as a level of language and discourse showed clear distinctions among essays included in both the independent and integrated datasets. This study also yielded results that confirmed McNamara, McCarthy and Crossley’s findings in terms of the CELEX logarithm for all words, mean. The descriptive statistics for this index showed results more favorable for integrated essays.

## **3.2.2 Syntactic complexity**

Most important, however, this study has immediate transfer potential in that it provides the indices that become an alert to assess the quality of essays. The principal component syntactic complexity required a qualitative analysis of the essays, mainly in the independent essay dataset in which some essays yielded indices above 8 words before the main verb, for example essay 1 in the independent essay dataset and essay 4 in the integrated one showed an index for SYNLE equal 10 which were very high. In both cases, the writer used a much-worded subject to introduce the topic sentence of the central paragraphs. For instance, in the following two examples readers can notice a sign of wordiness in the formulation of the topic sentence: “***One reason why online classes were not a good idea*** *is because not all the students had the necessary technology like computers, modern cellphones, cameras and also internet connection*”*.* Another example reads, “***One way that dating violence in teenagers can be manifested*** *is by being emotionally abused”.*

Panteleeva (2018) outlined that absolute values of syntactic complexity had better differentiate the worst essays from the best ones. The cluster of the criteria with absolute values includes the number of tokens, the maximum and minimum depth of sentences, the number of relative, adverbial and adjectival clauses and complex T-units. Panteleeva proposed the depth of sentence absolute value as the distance from the root of the tree to its farthest descendant. For example, the depth of sentence (1) is 5 (the path from the root America to the node than (Panteleeva, Lyashevskaya, & Vinogradova, 2018).Obviously, Pantieleeva’s proposal is very similar to SYNLE Coh-Metrix index proposed in this study.

*The most attractive for tourists in 1995 was North America, which was visited by more than 70 million people*.



Graph 7 Depth of sentence (Panteleeva, Lyashevskaya, & Vinogradova, 2018)

Another index that helps to distinguish between low and high quality essay is z-score syntactic simplicity. This measure indicates clearly that if the essay is above 0 z-score, this essay needs special attention and to be assessed its syntactic complexity. In graph 3 readers can notice several essays above 0 z-score. Graesser and et al. (2011) sustained that scores are higher when sentences have fewer words and simpler, more familiar syntactic structures. Contrariwise, if there are structurally embedded sentences that require the reader to hold many words and ideas in working memory, this implies large decreases of z-score, generally below 0 for college students (Graesser, McNamara, & Kulikowich, Coh-Metrix: Providing Multilevel Analyses of Text Characteristics, 2011).

## **3.2.3 The explicit textbase**

The analysis of two principal components of the explicit textbase level has also demonstrated how certain indices of referential cohesion and content word overlap influence on the quality of the essays that learners write. In the case of referential cohesion, both datasets showed moderate dispersion of the mean value, even though the independent essays dataset scattered a bit more as can be seen in the table 3 above. The descriptive statistics for content word overlap index showed that the integrated essay dataset showed better results than the independent essay dataset as illustrated in table 4.

Another index for the explicit textbase level that made distinction between high and low quality essay was lexical diversity. This measure of textual lexical diversity (MTLD) indicated that the integrated essay dataset was less scattered than the independent essay dataset. However, the independent essay dataset showed 51.21% essays above the mean value=92,43.

Crossley and et al. (2008) insisted on the importance of Coh-Metrix L2 reading index (Crossley, Greenfield, & McNamara, 2008). Thus, we concluded that if writing trainees were aware of the potential of certain Coh-Metrix variables that can improve the prediction of text readability, they would be, for sure, able to write essays for expected skilled L2 readers. These variables are related to lexical frequency, syntactic similarity, and content word overlap. The significance of these variables is that they broadly correspond respectively to three operations, which many psycholinguistic models of reading and comprehension distinguish: decoding, syntactic parsing, and meaning construction. These variables make up Coh-Metrix L2 reading index to better assess text readability.

As in all studies, this one has limitations. First, probably the students reading additional supporting materials before they started to write could have biased the independent essay dataset. Second, although the emphasis was not placed in Coh-Metrix L2 reading index, the researcher checked the three principal components with the RL2 indices in both datasets. However, Crossley and et al. stated that still this readability formula has been critiqued for lack of construct validity.

# 4. Conclusions

From a practical perspective, the findings of this study provide evidence for teachers to assume a new way to assess written production by considering three essential Coh-Metrix indices. These indices have proved to make distinctions between low and high quality essays. These indices are CELEX logarithm for all words, syntactic complexity expressed in words before the main verb, mean, and Measure of Textual Lexical Diversity (MTLD). In addition, other principal components indices of the multilevel framework of text characteristics have been analyzed: syntactic simplicity z-score, and content word overlap. It is also important to highlight that integrated writing tasks should be placed more emphasis on in our daily pedagogical practice. The descriptive statistics showed that integrated essays yielded better written production evidence than independent ones, in terms of the three principal components of the multilevel framework of text and discourse comprehension.

# 5. Bibliographic references

1. Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing Text Readability Using Cognitively Based Indices. *TESOL Quarterly*, 480- 490. doi:10.1002/j.1545-7249.2008.tb00142.x
2. Graesser, A. C., Cai, L. Z., & M., D. F. (2006). Question Understanding Aid (QUAID): A Web facility that helps survey methodologists improve the comprehensibility of questions. *Public Opinion Quarterly*, 3–22. .
3. Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, 231. doi:10.3102/0013189X11413260
4. Guoa, L., Crossley, S. A., & McNamara, D. S. (2013). Predicting human judgement of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing*, 233-234. doi:http://dx.doi.org/10.1016/j.asw.2013.05.002
5. McNamara, D., Mccarthy, P. M., & Crossley, S. A. (2010). Linguistic Features of Writing Quality. *Written Communication*, 58-76. doi:10.1177/0741088309351547
6. Panteleeva, I. M., Lyashevskaya, O. N., & Vinogradova, O. I. (2018). Inspector: The Tool for Automated Assessment of Learner Text Complexity. *National Research University Higher School of Economics (HSE).*, (pp. 9-11).