

Volume 5, Issue 2, 052505 (199-210)

Journal of Applied Linguistics Studies (JALS)

<https://doi.org/10.57647/jals.2025.0502.05>



The Effects of Writing Prompt Properties on Pre-Service Teachers' Essay Writing Performance Based on Coh-Metrix Analysis

Nader Afshounpour , Peyman Rajabi * , Hossein Ahmadi 

¹ Department of English Language, Mal. C., Islamic Azad University, Malayer, Iran

*Corresponding author: peymanrajabi25@malayeriau.ac.ir

Original Article

Received:
2025-06-10

Revised:
2025-07-30

Accepted:
2025-10-20

Published in Issue:
2025-12-30

©2025 The Author(s). Published by the OICC Press under the terms of the CC BY 4.0, [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract:

Writing prompts are widely used in second language writing assessment and instruction; however, the characteristics of these prompts are often treated as neutral elicitation tools. The properties of writing prompts may systematically influence how writers generate and organize their texts. This study investigated the effect of the writing prompt characteristics on Iranian pre-service English language teachers' essay writing. For this purpose, the research emphasized the evaluation of essays with six writing prompts with different properties. Sixty participants out of ninety randomly selected from the male population of Iranian EFL preservice teachers majoring in English at Farhangian University. They were at the age range of 20 to 24. All of them had prior experience in writing academic English paragraphs. They were taken writing courses to enhance their English writing ability, and they had formal training in English writing skills. This research was conducted both quantitatively and qualitatively. The quantitative analysis was incorporated to provide evidence as to whether the writing prompts elicit written essays that is stable and consistent across various prompts. Pre-service teachers' essays were analyzed using Coh-metrix to provide information about essay properties at different levels. For qualitative data, the interview recordings were transcribed and analyzed. Then common themes were extracted concerning essay composing. The findings suggest that writing prompts are not neutral instruments but important task conditions that shape the linguistic characteristics of 12 written texts. The results of the study showed that the properties of the prompt have an important effect on some textual characteristics of the response. For example, syntactic complexity, lexical complexity, and connectives. The basic conclusion of this research was that a range of prompt properties should be controlled to keep the prompts equal. Otherwise, the essays textual features changes because of the prompt. Integrating computational text indices with qualitative process data provides a richer understanding of how task conditions influence second language writing performance.

Keywords: Coh-Metrix, EFL Writing, Textual Features, Writing Prompt.

Cite this article: Afshounpour, N., Rajabi, P., Ahmadi, H. (2025) The Effects of Writing Prompt Properties on Pre-service Teachers' Essay Writing Performance based on Coh-metrix Analysis. *Journal of Applied Linguistics Studies*, 5(2), 199-210. <https://doi.org/10.57647/jals.2025.0502.05>

INTRODUCTION

Writing is the last skill we learned and the last skill we ever knew. Keys to good professional development writing is a significant challenge for students and crucial to success in business and career. For adults, writing skills are one of the best predictors of success in the first year of college

(Geiser & Studley, 2001). For professionals, writing skills are essential to their daily work and important for entry and advancement in their careers. Writing provides the ability to express ideas, discuss ideas, and connect multiple perspectives. Therefore, good writing is necessary to communicate with others, including teachers, colleagues, and the wider community (Crowhurst, 1990). Because

grammar is difficult to measure, writing is often only indirectly measured by grammatical knowledge. Thanks to developments in computer technology, we can now observe the properties of real samples. This tool not only allows us to evaluate students' writing in real time, but also tracks their progress, allowing English teachers/students to track their learning over time. Most people never consider the complexity and intricacy of the writing process. Ghaith (2002) in his article titled "The Nature of Writing Process" states that writing is a complex process that allows writers to work on ideas and make them visible and tangible. Writing stimulates thinking and learning because it encourages communication and creativity. Ideas can be reviewed, changed, added, and revised as they are written. Historically, writing testing relied on the difference between scores assigned to writing items to determine the correspondence of the response to the answer. In the past, the writing assessment field had relied on variance in scores awarded to written products to determine the effect of the prompt on the response (Hoetker, 1982; Hoetker & Brossell, 1989; Leu, Keech, Murphy, & Kinzer, 1982).

However, recently researchers have been able to use vocabulary and linguistic tools such as Coh-metrix and Corpus of Contemporary American English. This tool has been used to examine different types of writing in response to many varieties of prompts. There are very few studies exploring how different writing prompt properties affect the writing performance, especially Iranian preservice teachers. Modern computational linguistic tools are useful for investigating this issue. More recently, however researchers have been able to benefit from language corpora and computational linguistic tools such as Coh-metrix (Graesser, McNamara, Louwerse, & Cai, 2004) and the Corpus of Contemporary American English (Davies, 2008). Accordingly, this research can yield important implications for L2 writing classrooms to investigate the effects of different writing prompts properties on writing and seek to find new methods and techniques to teach writing in order to enable the students to write more meaningfully and more cohesively. Writing is a very important issue as academic institutions allocate a considerable amount of time to students' problems with writing performance. Therefore, the effect of prompt properties on writing should be investigated for this important issue. The impact of writing prompt properties on writing performance has been less investigated; therefore, studies are needed to investigate the interrelationship among prompt properties, writing performance and understanding of the help of a prompt as a scaffolding instrument. In other words, writing prompt may determine and shape the content of students' responses.

REVIEW OF THE RELATED LITERATURE

STUDIES RELATED PROMPTS

Kuiken & Vedder (2008) investigated how prompts of differing cognitive complexity affected a range of discourse measures. They administered two different prompts to 91

Dutch learners of Italian and 76 Dutch learners of French. The prompts were designed under a "complex condition" and a "non-complex condition" (Kuiken & Vedder, 2008: 52). Under the complex condition, the writers had to fulfill several more requirements to complete the task than under the non-complex condition. The responses were analyzed for accuracy (total number of errors and error severity), syntactic complexity, and lexical variation. The results indicated that fewer mistakes were made by both groups of language learners (French and Italian) in response to the complex condition than in response to the non-complex condition. There were significantly fewer total errors made and also significantly fewer severe errors in response to the complex condition. However, there were no significant differences in either syntactic complexity or lexical variation for either language group. That is, the written language elicited by the two prompt types was stable as measured by the syntactic complexity and lexical variation measures. The overall conclusion drawn by the authors was that increasing task complexity may elicit written language that is increasingly accurate for second language writers under exam conditions. Discourse analytic studies that have addressed prompt effect have drawn differing conclusions when investigating accuracy. While Kuiken & Vedder (2008), along with Way, Joiner, & Seaman (2000) observed variation in the accuracy of responses between different prompts, other studies (Zhang, 1987) found no significant differences in accuracy. The accuracy of second language writing is a key feature of the quality of the written product. Although prompt effect studies that have investigated the accuracy of responses are inconsistent in terms of the findings they have produced, the importance of accuracy to the construct or writing proficiency in a second or foreign language is so great that accuracy is a key writing competency that will be investigated in this study. Ong & Zhang (2010) administered writing prompts of varying task complexity to Chinese learners of English in Singapore. Task complexity was operationalized quite differently to the approach of other studies. The length of planning time before composing was varied, with three independent groups being allowed 20 minutes, 10 minutes or no planning time before beginning to write. The level of detail provided within the prompt was also divided into three distinct conditions, "(1) topic, ideas, and macro-structure given; (2) topic and ideas given; and (3) topic given", (Ong & Zhang, 2010). The final task complexity variable that was manipulated was the condition under which participants could revise their first drafts before submitting a completed piece of writing. There were only two conditions; revising with first draft visible and creating a final draft without being able to refer to the first draft. All three prompt variables were intended to provide differing levels of cognitive complexity. The responses were analyzed for fluency (total number of words per minute of writing and the mean number of words produced for each minute of total time spent on the task). Lexical complexity was operationalized as type-token ratio adjusted for text length. Ong & Zhang found no significant differences for the first

measure of fluency (total number of words per minute) for any task conditions. There was a significant difference in the second fluency measure (mean number of words per minute for total task time) based on the planning time task condition. Participants who did not have planning time produced significantly more words per minute on task than participants under the other planning conditions. Planning time also had a significant effect on lexical complexity. Again, responses produced without planning time were significantly more lexically complex than responses produced with either ten or twenty minutes planning time. Both of these findings are somewhat surprising as one may expect planning time to allow second language writers to prepare a lengthier response and to marshal their linguistic resources with less time pressure and hence make full use of their lexical resources. This was not the case in Ong & Zhang's study and the researchers speculated that the participants may not have utilized planning time effectively. Finally, Lim (2010) utilized a different approach to studying prompt effect to the work reported above. Lim developed a detailed prompt categorization framework and then used the Rasch model to create estimates of prompt difficulty based on the holistic scores awarded to MELAB essays. The difficulty of 60 distinct writing prompts that varied according to five key prompt characteristics were analyzed. Lim reported that the null hypothesis of all prompts being of equal difficulty was rejected, indicating that the prompts were not consistently equivalent, with a range of 2.78 logits between the easiest and the most difficult of the prompts (Lim, 2010). Lim stated that 14 of the 60 prompts analyzed should be removed from the operational prompt pool in order to establish a narrower prompt difficulty range (1.47 logits) and to better establish a claim of prompt equivalence. This provides some indication that there was variation in the equivalence of MELAB writing prompts analyzed that could potentially have contributed to differential test performance in terms of the score awarded. In terms of the prompt characteristics in Lim's categorization framework, the characteristic that exhibited most difference was topic domain. The largest differences were between prompts in the business and social domains). Lim concluded that scores awarded to the different prompts were stable apart from a few outlying prompts, with only topic domain having a potentially significant effect on scores. However, a careful reading of the results suggests that the detailed prompt categorization framework devised by Lim holds great promise for learning about prompt characteristics that contribute to prompt equivalence, or the lack thereof.

STUDIES RELATED TO WRITING DEVELOPMENT USING COH-METRIX

Kate Wolfe-Quintero et al. (1998, 4) analyzed the measures used in identifying development in L2 writing. They identified three categories related to different aspects of writing development: fluency, accuracy, and grammatical and lexical complexity. Recent advances in computational linguistics and discourse processing have shown possibilities that educators or researchers can automate

many language and text processing mechanisms. Several studies have employed automated text analysis methods for educational contexts especially in the area of writing assessment (Ullman, 2019) and discourse analysis (Ferretti & Lewis, 2019).

More than 50 published studies have demonstrated that Coh-Metrix indices can be used to detect subtle differences in text and discourse. Some of these studies used Coh-Metrix to distinguish different types of texts. For example, Louwse, McCarthy, McNamara, and Graesser (2004) identified significant differences between spoken and written samples of English. McCarthy, Lewis, Dufty, and McNamara (2006) reported that Coh-Metrix could successfully detect authorship even though individual authors recorded significant shifts in their writing style. Graesser, Jeon, Yang, and Cai (2007) identified differences between physics content that occurred in textbooks, texts prepared by researchers, and conversational discourse in tutorial dialogue. McCarthy, Briner, Rus, and McNamara (2007) showed that Coh-Metrix could differentiate sections in typical science texts, such as introductions, methods, results, and discussions. Lightman, McCarthy, Dufty, and McNamara (2007) distinguished the beginnings, middles, and ends of chapters in a corpus of history and science text books for high school. Hall, Lewis, McCarthy, Lee, and McNamara (2007) demonstrated that Coh-Metrix could distinguish between American-English law texts and British-English law texts. Crossley, Louwse, McCarthy, and McNamara's (2007) investigations of second language learner texts revealed a wide variety of structural and lexical differences between texts that were adopted (or authentic) versus adapted (or simplified) for second language learning purposes. Collectively, these studies demonstrate that Coh-Metrix is an extremely powerful text analysis tool, capable of assessing and differentiating an enormous variety of text types from the genre level to the sentence level.

Baba and Nitta (2010) conducted a study of EFL writing development in Japan using Coh-Metrix with a visual analysis of nonlinear individual development. They studied 46 students' writing on a repeated assignment over the course of one year. They found that their students improved in lexical diversity and grammatical complexity more than other text characteristics, such as speed and coherence. An analysis of the students' development as charted on line graphs showed that development patterns for the writing characteristics for most of the students were nonlinear. Their development patterns varied greatly. Baba and Nitta (2010, p.221) noted that "...even when the EFL students made some progress in writing they may not follow a sole, straight development path."

McNamara et al. (2010) studied 120 college students' essays graded by experts. The essays were untimed and written outside of class. They used Coh-Metrix to calculate scores on 53 indices. They found that only 3 indices predicted the quality of the essays: syntactic complexity, lexical diversity, and word frequency. They concluded that "...essay quality is judged largely by the sophistication of the writing rather than the ease of processing" (McNamara,

2010, 63). These results were confirmed in a later study (Crossley and McNamara, 2011 A) in which they reviewed a number of studies of essays written by first and second language writers. Here again, judgments of essay quality were best predicted by indices of language sophistication, such as more complex grammar and vocabulary.

Graesser et al. (2011, p.223) discuss the use of Coh-Metrix to identify "...major factors that account for most of the variance in texts across grade levels and text categories." They propose five text characteristics for readability discrimination. The first is narrativity which measures the extent to which the text tells a story. Narrativity is characterized by a higher percentage of verbs, adverbs, and intentional actions and events than informational passages, it tends to be about familiar topics, and it employs high frequency words and simple syntax (Graesser et al., 2011, p.231). The second factor is syntactic simplicity which is characterized by fewer words and simpler syntactic structures meaning shorter noun phrases. The third factor is word concreteness which means the content words are concrete and can more easily be visualized. The fourth factor is referential cohesion which means overlap in words and ideas in the passage. The final factor used in the computational analysis is causal cohesion which accounts for the use of causal and intentional connectives in the passage. From analyses of 37,520 texts of paragraph length covering K1 to 12th grade levels, they found that grade levels are primarily correlated with narrativity and syntax with only small variation in referential and causal cohesion across grade levels. Surprisingly, they found that in some cases words are more abstract at lower grade levels.

Crossley and McNamara (2011 B) studied characteristics of L2 writing across four different L1 backgrounds Czech, Finnish, German and Spanish and compared them to L1 writing. They found significant differences in Coh-Metrix indices for hypernymy, polysemy, stem overlap and lexical diversity. They found L2 writers used more general words with fewer semantic senses and that there was less cohesion as seen through less overlap in stems and less repetition of words. They suggested that L2 writers share some characteristics regardless of their first language due to the amount and type of linguistic knowledge they have. In a 2011 study, Lu evaluated 3,678 essays written by Chinese college students. Based on Wolfe et al.'s 1998 study, 14 measures of syntactic complexity were chosen to identify writing development using Lu's computational system. Seven measures were "...recommended to L2 writing researchers as good candidates for developmental indices" (Lu, 2011, p.56). The seven factors were as follows: complex nominals per clause, complex nominals per T-unit, length of clause, mean length of sentence, mean length of T- unit, and coordinate phrase per clause and per T- unit.

As seen from the research studies presented here, vocabulary, syntax, and cohesion are valuable aspects to be considered when evaluating L2 writing. In addition, recent focus in L2 development has pointed out the variability in L2 learners' development patterns. It is with this in mind

that the researcher has conducted an analysis of learners' paragraph writing both quantitatively and individually in an attempt to identify patterns of writing development.

For this purpose, the following research questions were asked.

1. What are the effects of writing prompt characteristics on EFL preservice teachers' essay writings in terms of linguistic features, i.e., Referential cohesion, Connectives, Syntactic complexity, Lexical diversity as detected by Coh-Metrix?
2. How do the prompt properties affect the Iranian EFL preservice teachers' essay composing?

INTRODUCING COH-METRIX INDICES

we present only a summary of Coh-Metrix's key indices. An extensive overview and analysis is provided in Graesser et al. (2004).

Causal Cohesion. Coh-Metrix calculates causal cohesion as the ratio of causal verbs to causal particles. Causal verbs such as kill, throw, and drop are identified through WordNet (Fellbaum 1998; Miller et al. 1990). Causal particles are identified in a pre-defined set and include items such as because, as a consequence, and the semantically depleted verbs make and cause.

Coreferential Cohesion. Referential links aid textual comprehension, facilitating inferencing, and benefiting recall (Kintsch and van Dijk 1978; McNamara 2001). Coh-Metrix employs four forms of lexical coreference identification: noun overlap, argument overlap, stem overlap, and (Latent Semantic Analysis) LSA-based semantic overlap. The overlap measures focus on comparing lexically based pairs such as table/tables and run/running. The LSA measures, on the other hand, employ singular value decomposition, a statistical technique, to analyze the semantic relationship between various textual elements. As such, LSA allows us to extend referential overlap beyond explicit relations such as chair/chairs into relative semantic similarities such as chair/table, table/wood, and wood/grass.

Connectives and Logical Operators. Connectives form cohesive links between separated sentential ideas. Coh-Metrix reports the density of connectives in various ways. For example, there are scores for positive-additive connectives (e.g., also moreover), negative-additive connectives (e.g., however, but), positive-temporal connectives (e.g., after, before), and negative-temporal connectives (e.g., until). Connectives serve as an extremely important indication of cohesion in a text (Haliday & Hasan 1976; Louwse 2002; Graesser et al. 2004). In addition, scores reflecting the density of logical operators such as or, and, and not are also reported. High densities of such items in a text place a high demand on the working memory of the reader.

Density of Major Parts of Speech. Coh-Metrix reports the incidence scores for various parts of speech (POS) as defined by the Brill (1995) POS tagger. These parts include pronouns, nouns, verbs, adjectives, adverbs, cardinal numbers, determiners, and possessives. Density scores

help to detect textual difficulty with, for example, a higher proportion of pronouns generally leading to a greater cognitive strain on the reader caused by more referential bridging (Graesser et al. 2004).

Polysemy and Hypernymy. Coh-Metrix tracks the ambiguity and abstractness of a text by incorporating WordNet (Fellbaum 1998) to calculate values for lexical polysemy (number of senses) and hypernymy (number of levels in a conceptual, taxonomic hierarchy).

Syntactic Complexity. The measure of syntactic complexity assumes that sentences with embedded constituents are either structurally dense, syntactically ambiguous, or ungrammatical (Graesser et al. 2004).

Word Information and Frequency. Word information incorporates four matrices: familiarity, concreteness, imageability, and meaningfulness. Coh-Metrix derives scores for these aspects via the MRC Psycholinguistic database (Coltheart 1981). High frequency words are those that are used more often (either in speech or writing) and are therefore likely to be more easily understood and read faster (Haberlandt & Graesser 1985; Just & Carpenter 1980).

Lexical diversity:

The measures of lexical diversity are on the number of words, the proportion of content words and Type-Token ratio. These measures were selected because lexical diversity is an indication of sophisticated writing skill

METHODOLOGY

PARTICIPANTS

This study was conducted at Khouzestan Farhangian University in Ahvaz, Iran. Sixty participants out of ninety were randomly selected from the male population of Iranian EFL preservice teachers majoring in English at Farhangian University. Participants were between 20 and 24 years old. All of them had prior experience in writing academic English paragraphs. They had taken writing courses to enhance their English writing ability, and they had formal training in English writing skills.

RESEARCH INSTRUMENTS

Coh-Metrix is an analysis tool that provides a variety of computational linguistic indexes related to vocabulary, syntax and semantics in order to satisfy the need for automatic and wide-range analyses. Coh-Metrix has been used to answer various research questions concerning distribution of linguistic features across text types, differences between argumentative and technical writing and difference in writing styles. In this study, we used Coh-Metrix 3.0 to analyze data for the various linguistic features of drafts comprehensively. In order to learn more about the effect of the writing prompt, the participants were interviewed, using a stimulated recall approach after they completed a writing test.

DATA COLLECTION

To serve the purpose of the present study, both qualitative and quantitative data were collected. At the initial stage (screening), 40 prompts were analyzed based on responses

from 30 students with the same level as trial writing prompt. Responses to the trial writing prompts were reviewed qualitatively and the decision to approve, revise or reject the trial prompt was made based on reading of responses and a review of feedback from the participants and raters. 6 prompts were selected based on distinguishing properties. According to Lim (2010), these properties included domain (educational, occupational, public, or personal), response mode (narrative or argumentative response), number of rhetorical cues (an instruction or question within the prompt that the writer must respond to) and open or focused (with no contextualization or with contextualization). One writing prompt was administered per session. All the essays were scored each session. Coh-metrix, an automated computational web tool was selected to analyze textual features. Each writing task was allotted 40 minutes. The quantitative approach was complemented by a qualitative approach to further investigate the effect of writing prompt properties on test takers' composing. Immediately after each session a stimulated recall interview was incorporated to recognize efficacy of writing prompt aspects on composing the response. For qualitative data, the interview recordings were transcribed and analyzed. Then common themes were extracted in relation to composing the essays. It should be mentioned that only narrative and argumentative types of prompts were used in this study.

DATA ANALYSIS

The quantitative analysis was conducted to determine whether written essays are permanent across various prompts. The MANOVA analysis was used to investigate the effects of writing prompt features on lexical diversity, cohesion, connectives and syntactic complexity. **Coh-metrix:** is an automated tool that provides linguistic indices for text and discourse (Graesser, McNamara, Louwerse, & Cai, 2004). Coh-metrix, an automated computational web tool will be selected to analyze textual features. Once the essays had been collected, all essays (N=360) were processed through the Coh-Metrix on four measures. Coh-metrix, an electronic correlation analysis tool, was selected for text analysis. The quantitative approach was accompanied by a qualitative approach to further investigate the effect of writing prompt properties on composing essays. In order to learn more about the effect of the writing prompt, preservice teachers were interviewed, using a stimulated recall approach after finishing each writing test. The purpose of these interviews was to participants recall how they interact with the writing prompt as they responded to and then composed a response under test conditions. The aim of the qualitative study was to recognize the aspects of writing prompt that may have effect on composing an essay.

RESULTS

The first research question was: What are the effects of writing prompt characteristics on EFL preservice teachers' essay writings in terms of linguistic features, i.e., Referential cohesion, Connectives, Syntactic complexity, Lexical diversity as detected by Coh-Metrix? The scores

obtained by the students were computed by Coh-Metrix to analyze their texts.

Table 1. shows the students' mean scores of three criteria of coherence (referential coherence and connections), syntactic complexity and lexical diversity. The first measure of the students' texts is referential cohesion with the mean of 0.37 which is below half. This mean is the total mean of all referential cohesion indices. Other total mean scores for other measurements, i.e., connectives, syntactic complexity and lexical diversity are 34.13, 1.25 and 37.74, respectively.

To study the organizational features of the students' essays, the following measures were analyzed. In this section, the findings of the study are illustrated one by one.

Reference: The first component of text analysis is referential cohesion. Coh-Metrix provides three general types of coreference indices. Noun overlap is overlap between nouns, with no deviation in form. Argument overlap is overlap between the noun in the target sentence and the same noun in singular or plural form in the previous sentence. Stem overlap is overlap from the noun to stems, regardless of word type (e.g., noun, verb, adjective). Thus, stem overlap could include overlap between *e.g. giver* in the target sentence and *giver, giving, or gave* in previous sentences. Both argument and stem overlap also include overlap between a pronoun and the same pronoun. Coreference indices also vary by distance between the target sentence and coreferent sentences. *Adjacent* overlap includes only adjacent sentences. Distances of *two* or *three* sentences include the target sentence and the two or three previous sentences, respectively. *All distances* consist of the overlap between each sentence with all other sentences in the text – this is intended as a more global index of cohesion.

Table 2 and Figure 1 below show the referential cohesion analysis of the students' passages.

This table shows the maximum and minimum for each index of this measure. It also shows the mean score and standard deviation for each index. According to this table, we can say that the maximum mean belongs to RC2 (Argument overlap, adjacent sentences) with the mean score of 0.619 and Std. deviation of 0.155 and minimum mean belongs to RC8 (Content word overlap, adjacent sentences, proportional) with the mean score of .105 and Std. deviation of 0.024.

Connectors: The second part of analyzing text is connectors. Another meaning of textual integrity is by connection. Conjunctions provide a clear indication of the type of relationship between ideas in a text and thus

increase the coherence of the text (Lewers, 2001). Coh-Metrix provides a frequency score (counts per 1000 words) for each link for different link types. The indicator shows five general categories of communication (Halliday and Hasan, 1976; Louwerse, 2001): relationship (because, therefore), relationship (and, or), inequality/unequal (although at the same time), physical relationships (first, up)) and additives (among others). There is also a difference between positive associations (also, plus) and negative associations (however, but).

Table 3. displays the descriptive statistics of “connective” for the studied corpus. This table shows the maximum and minimum for each index of this measure. It also shows the mean score and standard deviation for each index. With respect to this table, we can say that the maximum mean belongs to C11 (All connectives incidence) with the mean score of 109.41 and Std. deviation of 17.47 and minimum mean belongs to C14 (Adversative and contrastive connectives incidence) with the mean score of 16.1 and Std. deviation of 7.06.

Figure 2 gives a visual representation of the results.

Syntactic Complexity: The third component of text analysis is syntactic complexity. Coh-Metrix includes four methods. First, there is an index that calculates the average number of words that precede the main topic; assuming that the more words that precede the main subject, the more complex the syntactic structure will be. Second, there is an index that measures the average number of higher units per word (phrases and sentences), since higher units per word lead to more complex patterns. Coh-Metrix also evaluates an index that evaluates syntactic consistency by evaluating the similarity and consistency of syntactic patterns in the text. This measure not only works on the similarity of sentences, but also takes into account the speech in question. Grammatical consistency between sentences causes overlap between sentences, making the text easier to read. In contrast, sentences that exhibit compound structures (i.e., compound) are more difficult to process and understand (Perfetti, Landi, & Oakhill, 2005). Table 3. indicates the syntactic complexity of the studied corpora. This table shows maximum and minimum for each index of this measure. And also shows mean score and standard deviation for each index. With respect to this table, we can say that the maximum mean belongs to SC20 (Left embeddedness, words before main verb) with the mean score of 5.44 and Std. deviation of 2.21 and minimum mean belongs to SC26 (Similarity of sentence syntax, all combinations, across paragraphs) with the

Table 1. Mean Scores for the Students' Compositions

Measure	Mean
Referential Cohesion	0.37
Connectives	34.13
Syntactic Complexity	1.25
Lexical Diversity	37.74

Table 2. Descriptive Statistics for Referential Cohesion

	Minimum	Maximum	Std. Deviation	Mean
RC1	.147	.875	.167005	.48469
RC2	.251	.875	.155382	.61991
RC3	.284	.919	.170372	.56656
RC4	.163	.867	.165664	.44650
RC5	.290	.950	.150770	.58670
RC6	.229	.929	.173067	.53110
RC7	.041	.233	.045260	.12867
RC8	.058	.154	.024346	.10590
RC9	.038	.229	.043073	.11712
RC10	.060	.162	.020730	.10680

Note: RC=Referential Cohesion

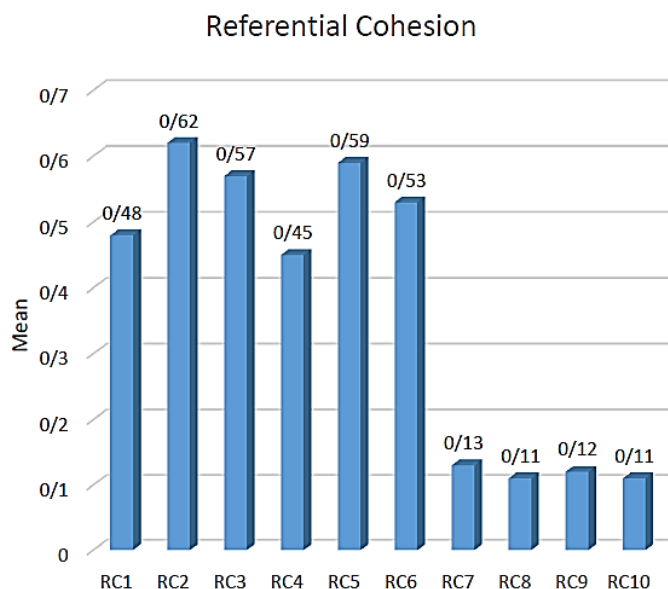


Figure 1. Referential cohesion indices of the EFL teachers

mean score of 0.079 and Std. deviation of 0.02.

Lexical Diversity: The fourth aspect of text analysis is lexical diversity. It refers to the different types (types) of words that appear by words (tokens) in the text. When the number of words is equal to all words (tags), each word is different. In this case, lexical diversity is maximum and the text is inconsistent or too short. When more words are used more frequently in the text, there will be fewer different words (and more cohesion). Coh-Metrix includes three different metrics: token type ratio (TTR), vod, and a measure of lexical text diversity (MTLD). TTR is the number of unique words divided by the number of word tags. The index generated by VOCD is calculated by comparing random TTR samples to the ideal TTR curve. MTLD is measured as the average length of a consecutive sentence in the text carrying the TTR value. TTR is relevant for long texts because as the number of single word tokens increases, the uniqueness

of these words decreases. Measurements such as VOCD and MTLD overcome this stress by using standardized models (McCarthy and Jarvis, 2010).

Table 5 shows the maximum and minimum for every index of this measure. It also shows the mean score and standard deviation for every index. With respect to this table, we can say that the maximum mean belongs to LD30 (Lexical diversity, VOCD, all words) with the mean score of 75.85 and Std. deviation of 16.21 and minimum mean belongs to LD28 (Lexical diversity, type-token ratio, all words) with the mean score of 0.52 and Std. deviation of 0.065.

As displayed in Table 6, MANOVA results revealed that there were significant effects of the writing prompt properties on lexical diversity, connectives and syntactic complexity. The results showed that the six writing prompts studied elicit essay writings that differ significantly in terms of a range of textual features. There are differences

Table 3. Descriptive Reports for Connectives Measurement

	Minimum	Maximum	Std. Deviation	Mean
C11	69.544	140.795	17.471071	109.41708
C12	22.207	62.668	9.590760	38.64045
C13	26.577	77.892	11.998773	47.49674
C14	2.825	30.639	7.060803	16.10567
C15	4.425	44.991	9.083811	16.84363
C16	5.553	38.628	8.025636	20.86222
C17	25.861	89.190	14.048191	57.84303
C18	.000	.000	.000000	.00000
C19	.000	.000	.000000	.00000

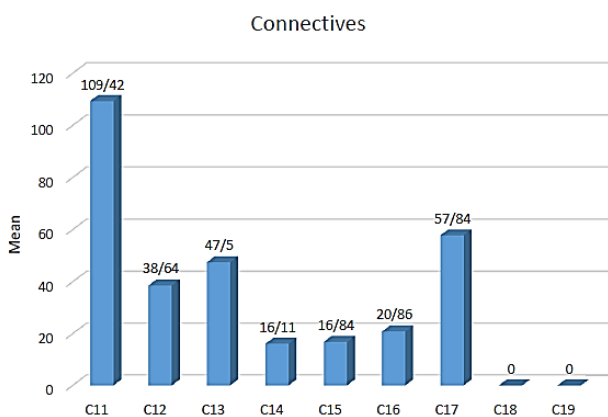


Figure 2. Connectives indices of the EFL teachers

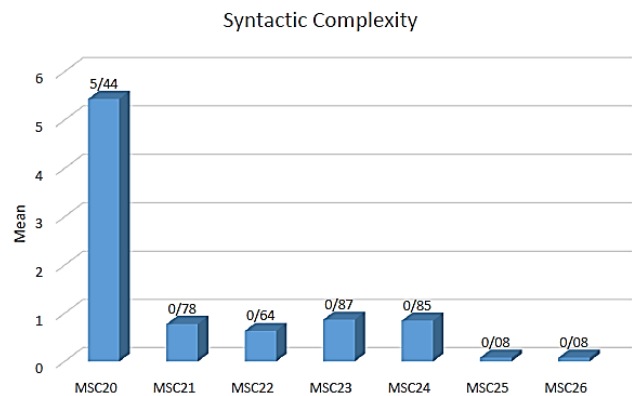


Figure 3. Syntactic complexity indices of the EFL teachers

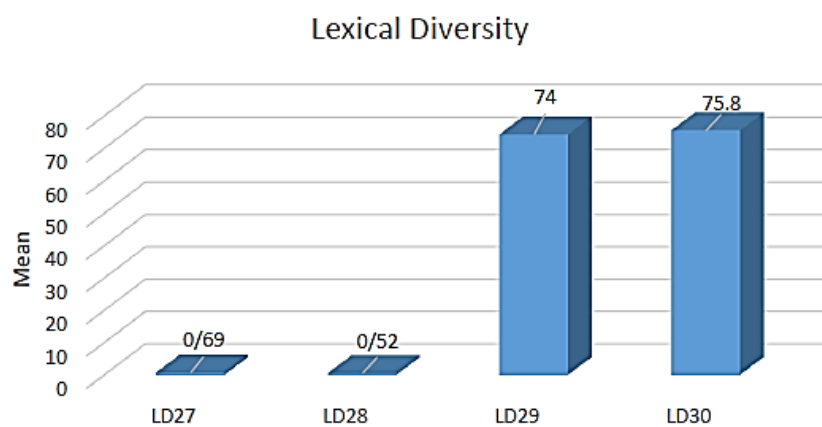


Figure 4. Lexical diversity indices of the EFL teachers

in the textual features based on given prompt such as lexical diversity, cohesion, connectives and syntactic complexity. The most significant effect was related to lexical diversity, connectives, and syntactic complexity.

DISCUSSION

To answer the first research question, a classification of prompt (Lim, 2010) was used to 40 writing prompts. In addition, experienced essay raters were asked to read

responses to 40 prompts that had been identified as eliciting different writing essays. Then 6 prompts were selected as the most distinguishing prompts. After that preservice teachers were interviewed to declare the prompt properties that were seen as important during the essay composing. According to Lim (2010) The distinguishing traits of writing prompts are:

- Domain (personal, public, occupational, educational)
- Response mode (narrative or argumentative)

Table 4. Descriptive Reports for Syntactic Complexity Measurement

	Minimum	Maximum	Std. Deviation	Mean
SC20	2.282	14.146	2.212108	5.44166
SC21	.563	1.040	.130195	.78460
SC22	.462	.723	.050536	.64490
SC23	.560	.939	.061518	.87476
SC24	.532	.920	.062740	.85457
SC25	.037	.118	.022225	.08080
SC26	.032	.128	.020155	.07950

Note: SC= Syntactic Complexity

Table 5. Descriptive Reports for Lexical Diversity Measurement

	Minimum	Maximum	Std. Deviation	Mean
LD27	.554	.877	.073266	.68794
LD28	.385	.678	.065730	.52358
LD29	46.294	120.963	17.160970	73.90170
LD30	42.502	122.604	16.217625	75.85772

Table 6. MANOVA Analysis of prompt effects

Factor score	Type III Sum of Squares	Df	Mean Square	F	Sig	Partial Eta Squared
Connectives	51.547	5	10.135	12.548	.000	.141
Syntactic Complexity	25.388	5	4.441	5.275	.000	.064
Referential Cohesion	2.659	5	.671	.670	.258	.008
Lexical Diversity	127.347	5	24.286	38.802	.000	.297

- Number of rhetorical cues
- Focus (open or focused)

Each session 60 essays were collected in response to each of six assigned writing prompts. These prompts differed based on the mentioned prompt properties. The 360 essays were analyzed in terms of important properties of foreign language writing proficiency such as syntactic complexity, lexical sophistication, cohesion and connectives. Coh-metrix and MANOVA analyses were utilized to measure the essays. There were significant effects of the writing prompt properties on lexical sophistication, connectives and syntactic complexity. The results reported in result chapter showed that the six writing prompts studied elicit essay writings that differ significantly in terms of a range of textual features. There are differences in the textual features based on given prompt such as lexical diversity, cohesion, connectives and syntactic complexity. The most significant effect was related to lexical sophistication,

connectives, and syntactic complexity. The results indicated that there are some clear relationships between prompt properties and the textual features of the essays written by EFL preservice teachers. In line with the current study Cho (2003) pointed out the need for test writers to be aware of the effects of writing prompts when they design them for assessment since they interact with other factors such as examinees' background and raters' scoring decisions in addition to determining the content of the response (p. 168). The results showed that the prompt properties of domain and mode of response most vividly elicit responses with significant differences in textual features. It is also clear that the prompt properties of focus and number of rhetorical cues elicit responses with significant differences in textual features. Two prompt properties that are vividly related with significantly different essay writings are domain and response mode. Essays written based on prompts placed in the personal domain that elicit a

narrative response are determined by high levels of lexical sophistication and low levels of syntactic complexity. Essays written based on prompts that are in non-personal domains and that elicit an argumentative response are identified by low levels of lexical sophistication and high levels of syntactic complexity. Contrary to the present study, Yu (2007) found that impersonal topics resulted in higher lexical diversity and Spaan (1993) speculated that second language writers use more multi-syllabic words in argumentative /impersonal composition. The main aim of the study was to investigate the prompt properties effects on essay writing performance summarized in the following table.

The above table is related to the first research question and shows that how the prompt properties affect the writing performance. Then concerning the second research question, the preservice teachers performed a stimulated recall interview. The interviews emphasized how preservice teachers compose an essay. The interview recordings were transcribed and analyzed. Common themes in relation to composing essays were extracted. The prompt properties of domain, number of rhetorical cues, and focus of the prompt were demonstrated to have an effect on participants' composing of the essays based on the prompt. Preservice teachers stated that the following prompt properties led to prompts that were relatively easy to write essays: prompts that were based on familiar topics, normally those placed in personal or occupational domains, prompts that had an open focus and gave them opportunity to respond on a wide range of various topics and prompts that had a large number of rhetorical cues because they provide structure for them. The recordings of the interviews revealed that the preservice teachers referred back to prompts while composing essays. Totally, three of the basic prompt properties (domain, focus, and number of rhetorical cues) had a range of effects on composing essays. Essays written to some prompts may be lexically rich but syntactically simple. However, essays written to other prompts may be syntactically complex and lexically simple. The recall interviews with the preservice teachers indicated that specific prompt properties had an effect on the interviewees' composing. The interviewees

emphasized that they are interested in prompts with familiar topics than other prompts. The interviewees wanted prompts that they can write long essays and perform better in terms of grade awarded based on a personally familiar topic. They also mentioned that they want a topic that allowed them to reflect the full domain of their lexical resources. In their views, prompts in personal domain were particularly easier to respond than prompts in other domains. In line with the present study Hoetker (1982) found that students perform better on the personal experience topic than the topic that required abstract reasoning. A number of studies investigated performance on prompts that invited a personal, first person response versus those that called for impersonal, third person responses (Brossell & Ash, 1984; Greenberg, 1981; Hoetker & Brossell, 1989). These studies found no significant differences. The other prompt property the preservice teachers were interested in, was prompts with a large number of rhetorical cues. Prompts with a large number of rhetorical cues were seemed easier than other prompts because the cues prepare the structure for the composing without the writers having to organize their own view. Because different prompts with different properties elicit written essays with different textual features, the prompts cannot be considered the same. The findings of the present study are in disagreement with Lim (2010). Lim concluded that there were no significant differences in scores awarded to responses in expository, argumentative, or narrative modes. According to the findings of the present study, the prompt properties have a significant effect on the textual features of the essays. Through knowledge of writing prompt features, learners can become more aware of the importance of strategy use, and this may be effective in promoting learners' self-confidence and potential in writing. Future research can investigate the effects of prompts at other features of essays and different language proficiencies.

CONCLUSION AND IMPLICATION

Accordingly, this research can have important implications for L2 writing classrooms to investigate the effect of

Table 7. prompt properties and essays textual features relationships

Prompt properties	Textual features of the essays
Personal domain, narrative response	levels of lexical sophistication (high) levels of connectives (low) levels of syntactic complexity (low)
Personal domain, open focus	levels of lexical sophistication (high) levels of cohesion (low) levels of syntactic complexity (low)
Argumentative response, non-personal domain, focused	levels of syntactic complexity (high) levels of lexical sophistication (low)
Argumentative response, focused, academic subject matter in public or educational domains	levels of lexical diversity (high)

different writing prompts on learners' writing performance. Writing is a very important subject because academic institutions devote considerable time to writing and seek to find new ways and techniques to teach writing to enable students to write more meaningfully and coherently. In addition, English teachers try to find possible reasons for students' problems with writing performance. Therefore, the role and effect of prompt properties in writing should be investigated for this purpose. Most studies have focused on the relationship between prompt and overall quality. The effect of writing prompt characteristics on writing performance has been less investigated. Therefore, studies are needed to investigate the interrelationship between prompt characteristics and writing performance. This can contribute to the ongoing debate in foreign language writing about the relationship between these two elements. If a significant relationship is found between prompt learning and better writing performance, it can be proven that prompts play an important role in writing. In addition, it can affect the way English writing is understood. And it may help teach creative writing because it involves deep thinking, organization, and revision. Through this type of writing prompt features, learners can become more aware of the importance of strategy use, and this may be effective in building learners' self-confidence and potential in language learning contexts. In addition, the findings of the present study suggest that writing instructors should teach EFL writers to apply complex syntactic structures in their essays.

Authors' contributions

All authors have contributed equally to prepare the paper.

Availability of data and materials

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Conflict of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- Baba, K., & Nitta, R. (2010, May). Dynamic Effects of Task Type Practice on the Japanese EFL University Student's Writing: Text Analysis with Coh-Metrix. In Twenty-Third International FLAIRS Conference.
- Brossell, G., & Ash, B. H. (1984). An Experiment with the Wording of Essay Topics. *College Composition and Communication*, 35(4), 423-425. <https://doi.org/10.58680/ccc198414859>
- Cho, Y. (2003). Assessing Writing: Are we Bound by only One Method? *Assessing Writing*, 8, 165-191. [https://doi.org/10.1016/S1075-2935\(03\)00018-7](https://doi.org/10.1016/S1075-2935(03)00018-7)
- Crossley, S. A., & McNamara, D. S. (2011). Understanding Expert Ratings of Essay Quality: Coh-Metrix Analyses of First and Second Language Writing. *International Journal of Continuing Engineering Education and Life Long Learning*, 21(2-3), 170-191. <https://doi.org/10.1504/IJCEELL.2011.040197>
- Crossley, S. A., Louwse, M. M., McCarthy, P. M., & McNamara, D. S. (2007). A Linguistic Analysis of Simplified and Authentic Texts. *The Modern Language Journal*, 91(1), 15-30. <https://doi.org/10.1111/j.1540-4781.2007.00507.x>
- Crowhurst, M. (1990). Teaching and Learning the Writing of Persuasive/Argumentative Discourse. *Canadian Journal of Education/Revue Canadienne de L'éducation*, 348-359. <https://doi.org/10.2307/1495109>.
- Davies, M. (2008). The Corpus of Contemporary American English: 450 million words, 1990-present.
- Ferretti, R. P., & Lewis, W. E. (2019). Best Practices in Teaching Argumentative Writing. In S. Graham, C. A. MacArthur, & J. Fitzgerald (Eds.), *Best Practices in Writing Instruction*, (3rd ed., 135-161). New York: Guilford Press..
- Geiser, S., & Studley, R. (2001). UC and SAT: Predictive Validity and Differential Impact of the SAT I and SAT II at the University of California. Oakland: University of California. https://doi.org/10.1207/S15326977EA0801_01
- Geiser, S., & Studley, W. R. (2002). UC and the SAT: Predictive Validity and Differential Impact of the SAT I and SAT II at the University of California. *Educational Assessment*, 8(1), 1-26. <https://doi.org/10.1207/S15326977EA0801-01>
- Ghaith, G. (2002). Using Cooperative Learning to Facilitate Alternative Assessment. In *Forum* (Vol. 40, No. 3, 26-31). <http://exchanges.state.gov/forum/>.
- Graesser, A. C., Jeon, M., McDaniel, B., & Graesser, A. (2008). Survey Interviews with New Communication Technologies: Synthesis and Future Opportunities. *Envisioning the Survey Interview of the Future*, 267-284. <https://doi.org/10.1002/9780470183373.ch14>
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix Measures Text Characteristics at Multiple Levels of Language and Discourse. *The Elementary School Journal*, 115(2), 210-229. <https://doi.org/10.1086/678293>
- Graesser, A. C., McNamara, D. S., Louwse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of Text on Cohesion and Language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202. <https://doi.org/10.3758/BF03195564>
- Graesser, Arthur, C., McNamara, D.S., & M. Kullikowich., J. (2011). Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, Vol. 40 (5), 223-234. <https://doi.org/10.3102/0013189X11413260>
- Greenberg, K. (1981). The Effects of Variations in Essay Questions on the Writing Performance of CUNY Freshmen. New York: The City University of New York Instructional Resource Center.
- Hall, C., McCarthy, P. M., Lewis, G. A., Lee, D. S., & McNamara, D. S. (2007). Using Coh-Metrix to Assess Differences Between English Language Varieties.
- Hall, C., McCarthy, P. M., Lewis, G. A., Lee, D. S., & McNamara, D. S. (2007). Using Coh-Metrix to Assess Differences Between English Language Varieties.
- Halliday, M.A.K. and Hasan, R. (1976) *Cohesion in English*. Longman, London, England
- Hoetker, J. (1982). Essay Examination Topics and Students' Writing. *College Composition and Communication*, 33(4), 377-392. <https://doi.org/10.2307/357950>
- Hoetker, J., & Brossell, G. (1989). The Effects of Systematic Variations in Essay Topics on the Writing Performance of College Freshmen. *College Composition and Communication*, 40(4), 414-421. <https://doi.org/10.2307/358240>
- Just, M.A. & Carpenter, P.A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review*, 87, 329-354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Kuiken, F., & Vedder, I. (2008). Cognitive Task Complexity and Written Output in Italian and French as a Foreign Language. *Journal of Second Language Writing*, 17(1), 48-60. <https://doi.org/10.1016/j.jslw.2007.08.003>
- Leu, D. J., Keech, K. L., Murphy, S., & Kinzer, C. (1982). Effects of Two Versions of a Writing Prompt upon Holistic Score and Writing Processes. Gray, JR and Ruth, LP, *Properties of Writing tasks: A Study of Alternative Procedures for Holistic Writing Assessment*, Berkeley: University of California, Graduate School of Education, Bay Area Writing Project, 215-219.
- Lightman, E. J., McCarthy, P. M., Dufty, D. F., & McNamara, D. S. (2007). Using Computational Text Analysis Tools to Compare the Lyrics of Suicidal and Non-suicidal Songwriters. In *Proceedings*

- of the Annual Meeting of the Cognitive Science Society (Vol. 29, No. 29).
- Lim, G. S. (2010). Investigating Prompt Effects in Writing Performance Assessment. *Spain fellow Working Papers in Second or Foreign Language Assessment* ,8, 95-115.
- Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in Language and Cohesion Across Written and Spoken Registers. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 26, No. 26).
- McCarthy, P. M., Briner, S. W., Rus, V., & McNamara, D. S. (2007). Textual Signatures: Identifying Text-Types Using Latent Semantic Analysis to Measure the Cohesion of Text Structures (in *Natural Language Processing and Text Mining*).
- McCarthy, P. M., Lewis, G. A., Dufty, D. F., & McNamara, D. S. (2006, May). Analyzing Writing Styles with Coh-Metrix. In *Flairs Conference* (764-769).
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511894664>
- McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). Coh-Metrix: Automated Cohesion and Coherence scores to Predict Text Readability and Facilitate Comprehension. Technical report, Institute for Intelligent Systems, University of Memphis, Memphis, TN.
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing Linguistic Features of Cohesion. *Discourse Processes*, 47(4), 292-330. <https://doi.org/10.1080/01638530902959943>
- McNamara, D., Louwerse, M., & Graesser, A. (2002). Coh-Metrix: Automated Cohesion and Coherence Scores to Predict Text Readability and Facilitate Comprehension (Grant Proposal).
- McNamara, D.S., Crossley, A.S., & McCarthy, P.M. (2010). Linguistic Features of Writing Quality, *Written Communication*, Vol. 27 (1), 57-86. <https://doi.org/10.1177/0741088309351547>
- McNamara, D.S., Graesser, A.C., Cai, Zh., & Kulikowich, J.M. (2011). Coh-Metrix Easability Component: Aligning Text Difficulty with Theories of Text Comprehension. AERA 2011. Retrieved from the Internet 23 May 2012.
- McNamara, T. (2010). The Use of Language Tests in the Service of Policy: Issues of Validity. *Revue Française de Linguistique Appliquée*, 15(1), 007-023.
- McNamara, T. F. (1996). *1996: Measuring Second Language Performance*. London and New York: Addison Wesley Longman.
- Ong, J., & Zhang, L. J. (2010). Effects of Task Complexity on the Fluency and Lexical Complexity in EFL Students' Argumentative Writing. *Journal of Second Language Writing*, 19(4), 218-233. <https://doi.org/10.1016/j.jslw.2010.10.003>
- Spaan, M. (1993). The Effect of Prompt in Essay Examinations. In D. Douglas & C. Chapelle (Eds.), *A New Decade of Language Testing Research* (98-122). Alexandria, VA: TESOL.
- Ullmann, T. D. (2019). Automated Analysis of Reflection in Writing: Validating Machine Learning Approaches. *International Journal of Artificial Intelligence in Education*, 29, 217-257. <https://doi.org/10.1007/s40593-019-00174-2>.
- Way, D. P., Joiner, E. G., & Seaman, M. A. (2000). Writing in the Secondary Foreign Language Classroom: The Effects of Prompts and Tasks on Novice Learners of French. *The Modern Language Journal*, 84(2), 171-184. <https://doi.org/10.1111/0026-7902.00060>
- Wolfe-Quintero, Kate, Inagaki, Sh., & Kim., H. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity* (Technical Report). Manoa, Hawaii: Second Language Teaching and Curriculum Center, University of Hawaii.
- Yu, G. (2007). Lexical Diversity in MELAB Writing and Speaking Task Performances. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 5, 79-116.
- Zhang, S. (1987). Cognitive Complexity and Written Production in English as a Second Language. *Language Learning*, 37(4), 469-481. <https://doi.org/10.1111/j.1467-1770.1987.tb00580.x>