Journal of Contemporary Educational Research, 2024, Volume 8, Issue 1
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

A Study on the Correlation Between LIWC Word Categories and Chinese Composition Writing Performance of Fourth, Fifth, and Sixth Grade Students

Yufeng Wu*

Affiliated Foreign Language School of Longhua District Education Science Research Institute, Shenzhen 518000, Guangdong Province, China

***Corresponding author:** Yufeng Wu, 724129215@qq.com

Abstract: This study focuses on the analysis of the Chinese composition writing performance of fourth, fifth, and sixth grade students in 16 selected schools in Longhua District, Shenzhen during the spring semester of 2023. Using LIWC (Linguistic Inquiry and Word Count) as a text analysis tool, the study explores the impact of LIWC categories on writing performance which is scaled by score. The results show that the simple LIWC word categories have a significant positive influence on the composition scores of lower-grade students; while complex LIWC word categories have a significant negative influence on the composition scores of lower-grade students but a significant positive influence on the composition scores of higher-grade students. Process word categories have a positive influence on the composition scores of all three grades, but the impact of complex process word categories increases as the grade level rises.

Keywords: Chinese composition; LIWC word categories; Writing performance; Grades; Text mining

Online publication: January 20, 2024

1. Introduction

The study of factors influencing Chinese composition scores has always been a hot topic in the case of language teaching and studying. Current research on the factors influencing composition scores mainly focuses on second language learning, with an emphasis on three perspectives: learning psychology, writing strategies, and writing instruction. Research from a learning psychology perspective discusses the influence of student anxiety and self-efficacy on composition scores [1]. The writing strategies perspective focuses on factors during the writing process, such as the time spent on ideation [2] and the timing of composition submission [3] and their impact on composition scores. The writing instruction perspective considers the influence of different teaching methods on student composition scores, such as the effects of goal-oriented discussions and feedback for improvement [4], and the impact of metacognitive strategy training [5]. However, these studies lack attention to the text features of

compositions.

Regarding specific research on factors influencing Chinese composition scores, most studies start from the perspective of writing instruction. For example, some studies explore the impact of using new technologies [6], different teaching methods [7], and reflective writing instruction [8] on Chinese composition scores. There are also some quantitative studies that analyze the correlation between writing attitudes and composition scores using attitude scales [9] or investigate the relationship between anxiety levels and Chinese writing skills [10]. However, most of these studies aim to optimize teaching methods to improve Chinese composition scores, resulting in primarily experiential research on teaching methods. There are relatively fewer quantitative studies and even fewer quantitative studies based on text analysis.

Based on this, this study attempts to analyze the effects of text features, such as word types and different process words, on Chinese composition scores in fourth, fifth, and sixth grade students. This will be achieved using the text analysis software LIWC (Linguistic Inquiry and Word Count) to process the compositions and extract relevant text features.

LIWC is a language processing software that includes a built-in dictionary that categorizes words into various categories. It analyzes text by segmenting it into words and counting the frequency of word usage in each category [11]. LIWC has been widely used in linguistic analysis, such as analyzing the use of emotion words in text [12], comparing the lexical diversity between spoken and written texts using LIWC [13], and detecting depressive tendencies in college students' writing using LIWC [14]. The first version of LIWC was released in 1999, and the latest version was released in 2022. The words in the LIWC dictionary can be classified into two main categories: function words and process words.

Function words include pronouns, prepositions, articles, conjunctions, and modal verbs, which are semantically empty. Function words reflect the writer's expression style. For example, the frequent use of the first-person singular pronoun "I" reflects the writer's self-focus, while the frequent use of the first-person plural pronoun "we" reflects the writer's focus on the collective [11].

Process words reflect the psychological processes that the writer is concerned about, thinking about, avoiding, and the writer's psychological experiences, as well as how they organize and analyze the world. For example, emotion words that express the writer's emotions, such as "happy" or "angry," are categorized as emotional process words. Previous studies have confirmed a significant correlation between learners' emotional states and composition scores. For instance, anxiety and fear have a negative impact on composition scores [15], while an active emotional state has a positive impact on composition scores [16]. Since emotion words themselves express learners' emotional states, this study considers emotion words as a predictor of learners' composition scores, aiming to explore whether the use of emotion words contributes to better composition scores.

2. Research methodology

2.1. Text selection

For this study, student essays from the mid-term Chinese language assessment for grades four to six in selected primary schools in Longhua District, Shenzhen, in the spring of 2023 were chosen as the research texts. Although the assessment was organized independently by each school, the same essay prompts were used, making them suitable for research. The sample sizes were as follows: 4th grade (4049 essays), 5th grade (3735 essays), and 6th grade (3805 essays).

2.2. Instruments

This study utilized LIWC to analyze the essays and obtain the results of students' usage of LIWC word

categories. SPSS27 was used to analyze the correlation of LIWC word categories and students' writing performance which is scaled by scores.

2.3. Procedure

The student essays were scanned as images and then converted into text using an image-to-text conversion tool. After comparing various tools, WPS was chosen for its higher accuracy in text conversion. The resulting text was imported into the LIWC system for Chinese text analysis, obtaining various text indicators such as word categories and temporal word frequencies. Some word categories that do not apply to Chinese such as articles, words that indicate tense, and apostrophe were excluded.

Statistical analysis and processing were conducted using SPSS27. Based on the nature of the data, methods such as analysis of variance, *t*-tests, and correlation analysis were primarily employed.

3. Results

By conducting a correlation analysis between the LIWC word categories and essay scores, the following correlation patterns were observed.

3.1. The impact of word categories on essay grades

According to the results of Pearson correlation analysis shown in **Table 1**, the usage of adverbs, conjunctions, quantifiers, and numbers significantly influences the scores. Specifically, adverbs, conjunctions, and numbers have a strong impact on the scores. The usage of adverbs, conjunctions, and quantifiers positively affects the essay scores in all three grades, while the usage of numbers has a negative impact on the scores across all three grades. Adverbs modify verbs and using more adverbs indicates a higher level of language proficiency in providing sufficient modifications to verbs. Similarly, using a higher number of conjunctions implies the use of more complex sentences, which reflects more proficient language expression skills. Quantifiers belong to the more complex aspects of Chinese grammar, so using them more frequently indicates stronger language expression skills. On the contrary, using a higher number of numerals, which are grammatically simpler, negatively correlates with essay scores in all three grades. Using numerals excessively is often seen as an attempt to meet the word count, resulting in lower essay scores.

In addition to these consistent tendencies across grades, there are other text features whose influence varies with grade level. These text features can be categorized into three types.

The first type includes text features that do not significantly affect or have a negative impact on the essay scores in the fourth grade but have a relatively significant impact on the scores in the fifth and sixth grades. Examples of such features are function words, negations, and particles. Function words refer to words that serve grammatical functions without carrying substantial meaning, such as prepositions, articles, conjunctions, pronouns, and auxiliary verbs. The usage of function words shows a significant negative correlation with essay scores in the fourth grade but exhibits a significant positive correlation in the fifth and sixth grades (correlation coefficients of 0.111 and 0.158, respectively). For fourth grade learners, the accuracy of using function words is lower compared to the fifth and sixth grades. Therefore, the usage of function words negatively affects essay scores in the fourth grade but positively affects those in the fifth and sixth grades. Using a higher number of function words indicates a higher level of language expression skill when used correctly. Similarly, the usage of negations and particles does not have a significant impact on the essay scores in the fourth grade but shows a relatively significant positive impact on the scores in the fifth and sixth grades.

Table 1. The correlation between LIWC word categories and essay scores

Word categories	Forth	Fifth	Sixth
Adverb	0.128***	0.147***	0.174***
Conj	0.209***	0.12***	0.134***
Quant	0.041**	0.078***	0.041**
Number	-0.356***	-0.516***	-0.039**
Funct	-0.094***	0.111***	0.159***
Negate	0.003	0.04**	0.042**
Interjunction	-0.025	0.175***	0.103***
You	0.145***	0.036**	-0.01
Preps	0.082***	0.082***	0.022
Verb	0.099***	0.149***	0.002
AuxVerb	0.153***	0.118***	0.01
Pronoun	-0.066***	0.088***	-0.01
I	-0.135***	0.039**	-0.042**
YouPL	-0.015	0.008	-0.055**
SheHe	-0.045**	0.02	0.004
They	-0.024	0.021	0.034**
We	-0.021	0	-0.031

The second type includes text features that significantly affect the essay scores in the fourth and fifth grades but have no impact on the scores in the sixth grade. Examples of such features are second-person pronouns, prepositions, verbs, and auxiliary verbs. Second-person pronouns, prepositions, verbs, and auxiliary verbs are commonly used words. For fourth and fifth grade learners, using these words more frequently can reflect a stronger language expression skill to some extent. However, for sixth-grade students who already have more language experience, using these common words more frequently is considered a basic language expression skill and does not significantly impact essay scores.

The third type includes text features whose impact on essay scores varies in a complex manner across all three grades. Examples of such features are pronouns, first-person singular pronouns, second-person possessive pronouns, third-person singular pronouns, and third-person plural pronouns. For fourth grade essay scores, pronouns, first-person singular pronouns, and third-person singular pronouns have a significant negative impact. For fifth grade essay scores, pronouns and first-person singular pronouns have a significant positive impact. For sixth grade essay scores, first-person singular pronouns and second-person singular pronouns have a significant negative impact, while third-person singular pronouns have a significant positive impact. These text features are all pronouns, and the complex effects observed may be related to the specific characteristics of pronouns.

From the results, it can be concluded that although the difficulty of learning different pronouns varies, using a higher number of pronouns in the fourth grade often has a negative impact on learning outcomes. This may be because fourth grade students may not yet be able to use pronouns correctly, and they tend to excessively use first-person pronouns, as indicated by the correlation coefficient of -0.135. However, in the fifth grade, using various pronouns appropriately can have a positive impact on essay scores. For sixth grade students, using more third-person pronouns indicates a shift in the writing perspective from self to others, which signifies a higher level of expressive skill. Therefore, third-person pronouns have a positive impact on essay

scores, while first-person and second-person pronouns have a negative impact.

3.2. The impact of different process words on essay scores

LIWC categorizes words into five types of process words: social process words, emotional process words, cognitive process words, perceptual process words, and biological process words, totaling 28 text features. The correlation analysis results between these words and essay scores are shown in **Table 2**.

Table 2. The correlation between LIWC process words and essay scores

Word categories	Forth	Fifth	Sixth
Social	0.072***	0.1***	0.072***
Family	-0.017	0.04**	0.034**
Friend	-0.025	0.04**	0.017
Humans	-0.01	0.029	0.107***
Affect	0.097***	0.203***	0.205***
PosEmo	0.054**	0.152***	0.176***
NegEmo	0.07***	0.095***	0.047**
Anx	0.096***	0.059***	0.066***
Anger	0.02	0.049**	0.012
Sad	0.034	0.056**	0.036**
CogMech	0.235***	0.186***	0.147***
Insight	0.136***	0.117***	0.105***
Cause	0.166***	0.078***	0.094***
Discrep	0.148***	0.1***	-0.022
Tentat	0.087***	0.132***	0.022
Certain	-0.175***	0.099***	0.103***
Inhibition	0.049**	0.067**	0.071***
Inclusive	0.061***	0.056***	0.05**
Exclusive	0.146***	0.055***	0.036**
Percept	0.211***	0.167**	0.168***
See	0.155***	0.111**	0.138***
Hear	0.013	0.09***	0.049**
Feel	0.145***	0.098***	0.067**
Bio	0.094***	0.074***	0.05**
Body	0.089***	0.129***	0.071***
Health	0.035	0.055**	0.055**
Sexual	-0.011	0.017	0.037**
Ingest	0.042**	-0.015	0.022

Social process words have a significant positive impact on the Chinese composition scores of fourth, fifth, and sixth graders, but the correlation is weak (0.072, 0.1, and 0.072, respectively). Specifically, family words have a significant positive impact on the writing of fifth and sixth graders (0.04 and 0.034), friend words have a significant positive impact on the composition scores of fifth graders (0.04), and human words have a significant positive impact on the composition scores of sixth graders (0.107). However, the correlation coefficients of these impacts are relatively low.

Emotional process words are significantly positively correlated with the composition scores of fourth, fifth, and sixth graders. It is worth mentioning that although they are all significantly positively correlated, the impact coefficient of positive emotional words on composition scores is similar to negative emotional words at the beginning but becomes two to three times higher than that of negative emotional words as the students' grades increase. This indicates that the impact of positive emotional words on composition scores is increasing. Among the negative emotional words, only anxiety words have a significant positive impact on the composition scores of fourth, fifth, and sixth graders, while anger words and sadness words only have a significant positive impact on the composition scores of some grades. It can be concluded that although negative emotional words have a significant impact on composition scores, their correlation coefficients are not as strong as those of positive emotional words.

Perceptual process words have a relatively high impact on the composition scores of fourth, fifth, and sixth graders, with coefficients of 0.211, 0.167, and 0.168, respectively. Perceptual process words have the highest impact on the composition scores of fourth graders, but their impact decreases significantly for fifth and sixth graders. This indicates that fourth grade students who express their perceptions more in their writing tend to achieve higher composition scores. Specifically, the use of visual words has a stable level of impact on composition scores across fourth, fifth, and sixth grades, while auditory words and sensory words show a noticeable decrease in impact as the students' grades increase. Therefore, younger students who express their perceptual processes more are more likely to receive higher composition scores.

Physiological process words also have a significant positive impact on the composition scores of fourth, fifth, and sixth graders, but the correlation coefficients are all below 0.1. The impact of physiological process words on composition scores also decreases steadily as the students' grades increase. Specifically, the impact of body words on composition scores fluctuates without a specific pattern. On the other hand, health words, sex words, and eating words only have a significant impact on the composition scores of some grades.

Comparing the changing trends in the impact of social, emotional, cognitive, perceptual, and biological process words on the composition scores of fourth, fifth, and sixth graders, as shown in **Figure 1**, it can be concluded that the influence of emotional process words on composition scores increases with grade level. However, the impact of cognitive process words, perceptual process words, and physiological process words decreases with grade level. The impact of social process words increases from fourth to fifth grade and then decreases in sixth grade. Therefore, although all process words have a significant positive impact on composition scores, the influence of emotional process words continuously increases and surpasses the other process words in sixth grade, becoming the most influential factor. It can be inferred that as students progress to higher grade levels, expressing their emotions effectively will lead to better composition scores.

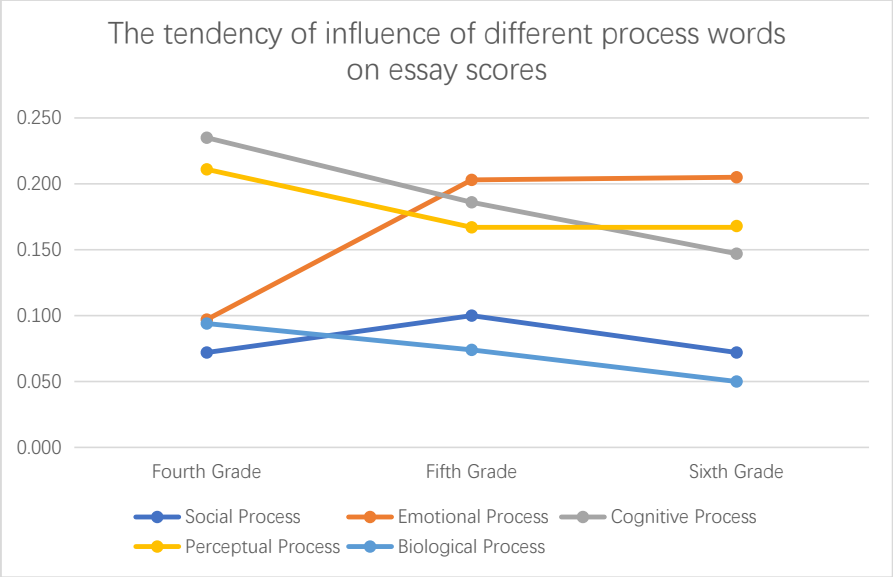

Figure 1. The tendency of influence of different process words on essay scores

4. Discussion

This study processed a sample of simulated compositions from selected schools in Longhua District, Shenzhen, using the Wenxin system and conducted a correlation analysis between various text features measured by the LIWC and the authors' scores. Three conclusions were drawn from the results.

Firstly, simple word categories have a noticeable positive impact on the composition scores of lower-grade students, but they do not have a significant effect on the scores of higher-grade students. Complex word categories have a significant negative impact on the composition scores of lower-grade students but have a significant positive impact on the scores of higher-grade students.

Secondly, complex word categories and process words have a negative impact on the composition scores of lower-grade students but a positive impact on the scores of higher-grade students. This is because lower-grade students have lower accuracy in using complex word categories and process words, and excessive use can lead to errors and lower composition scores.

Thirdly, all five types of process words have a significant positive impact on the composition scores of fourth, fifth, and sixth grade students. However, the impact of cognitive, perceptual, and emotional process words is significantly greater than that of social and physiological process words. Among them, only the impact of emotional process words increases with the grade level of the students.

This study examined the impact of different LIWC word categories on the composition scores of fourth, fifth, and sixth grade students. The text recognition process used the image-to-text conversion tool in WPS, which may have some errors. Therefore, using a more accurate text recognition tool could improve the accuracy of this study. Additionally, the study focused on fourth, fifth, and sixth grade students, and it may have yielded more findings if the grade range was broader.

Disclosure statement

The author declares no conflict of interest.

References

[1] Li H, Liu R, 2013, The Relationship Between Foreign Language Writing Anxiety and Writing Self-Efficacy of College Students and Its Prediction on Writing Achievement. Foreign Language Research, 2(7): 48–54.

[2] Dai J, Xu S, 2001, A Study on the Allocation of Pre-Composition Thinking Time and its Influence on Writing Achievement. Psychological Development and Education, 17(2): 36–40.

[3] Duan W, Year, A Study on the Correlation Between High School Students' English Writing Strategies and Writing Achievement, dissertation, Minnan Normal University, 2016.

[4] Li W, Li F, 2003, An Experimental Study on the Effects of Goal Orientation and Progress Feedback on Writing Achievement of Top and Bottom Students. Psychological Science, 26(6): 1086–1090.

[5] Xiao W, 2011, An Empirical Study on the Improvement of Students' English Writing Proficiency by Metacognitive Strategy Training. Journal of Foreign Languages, 27(5): 124–127.

[6] Zhao C, 2020, Application Strategies of Information Technology in Middle School Mathematics Teaching Classroom. Literature Youth, 2020(020): 1.

[7] Zhu W, 2017, The Application of Situational Teaching in Junior Middle School Chinese Writing Teaching. Urban Tutoring: The Second Half of the Month, 2017(6): 51.

[8] Duan X, 2014, Reflection on Chinese Writing Teaching in Senior High Schools. Reading and Writing: Journal of Education and Teaching, 11(11): 1.

[9] Li M, Li W, Bei W, 2021, A Study on the Correlation Between Chinese Writing Attitude and Writing Achievement of High School Students. Teaching Reference for Middle Schools, 2021(24): 3.

[10] Wang L, 2008, A Study on the Correlation Between Second Language Writing Anxiety and Native Language Writing Ability. Journal of Gansu Normal University, 13(3): 3.

[11] Tausczik YR, Pennebaker JW, 2010, The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology, 29(1): 24–54.

[12] Kahn JH, Tobin RM, Massey AE, et al., 2007, Measuring Emotional Expression with the Linguistic Inquiry and Word Count. The American Journal of Psychology, 120(2): 263–286.

[13] Yu G, 2010, Lexical Diversity in Writing and Speaking Task Performances. Applied Linguistics, 31(2): 236–259.

[14] Lumontod III RZ, 2020, Seeing the Invisible: Extracting Signs of Depression and Suicidal Ideation from College Students' Writing Using LIWC a Computerized Text Analysis. Int. J. Res. Stud. Educ, 2020(9): 31–44.

[15] Chandler S, 2007, Fear, Teaching Composition, and Students' Discursive Choices: Re-Thinking Connections Between Emotions and College Student Writing. Composition Studies, 35(2): 53–70.

[16] Shao KQ, Yu WH, Ji ZM, 2013, The Relationship Between EFL Students' Emotional Intelligence and Writing Achievement. Innovation in Language Learning & Teaching, 7(2): 107–124.