

# Comparative Analysis of the Readability of English Research Articles Written by Chinese and English Native Writers

Meiying Song, Qiwei Dai, and Hang Yu\*

**Abstract**—This research tries to explore the linguistic factors affecting the citation rate of Chinese writer’s research articles. We selected 50 SCI journal article introductions written by English Native and Chinese writers respectively. We measured their readability using Coh-Metrix. We then analyzed the data based on Graesser & McNamara’s theory of multi-level discourse comprehension. Our results show that the introductions written by the two types of writers differ significantly at the following levels: (1) at the discourse level, the referential cohesion score of the Chinese writer’s introductions was higher than that of the native speaker’s, indicating a higher readability; (2) at the situation model level, the deep cohesion and time sequential scores of the native writer’s introductions were higher than those of the Chinese writer’s, suggesting higher readability. Our findings also imply that although some readability indices of Chinese writers’ introductions are higher than those of the native writer’s, it does not necessarily indicate high readability literally, nor good writing quality, which could also be an explanation for the low citation rate. Moreover, the results show that the readability parameters of Chinese and native writer’s introductions differ at various levels, so it is hard to conclude that the readability alone impacts the citation rate.

**Index Terms**—Coh-Metrix, introduction, journal article, readability, citation rate

## I. INTRODUCTION

International publications represent the research capability and standard of a country, and also is an important measure for evaluating an individual’s research abilities and career advancement by universities and employers. Additionally, the publication of journal papers can enhance an individual’s confidence, improve one’s academic writing skills, and increase employability. Therefore, the demand for international publications in academia is increasing rapidly, and the competition is becoming more and more intense. As the primary medium for international publications, the use of English presents a significant challenge for scholars from non-English speaking countries.

To improve the quality of writing and increase international publication rates, various studies have been conducted by researchers in China on writing SCI journal articles [1–3]. In recent years, Chinese scholars have jumped to the world second in terms of the total number of papers published in international journals. However, the ultimate goal of the international publication is to enhance scientific research influence, and paper citation rate is a strong index reflecting such influence. However, although the total number of papers published by Chinese scholars in international journals has increased, their citation rate is relatively low [4]. According to the Report on Science Development published by the Chinese Academy of Sciences

in 2015, Chinese researchers published about 1.37 million papers in SCI journals, second only to the United States. However, the citation rate ranked the fourth and was far lower than that of Germany (about 900,000 papers) and the United Kingdom (about 800,000 papers), which ranked the third and fourth respectively in terms of the number of papers published. This has also drawn attention from the public: “why have Chinese scholars’ papers not received enough attention from the international academic community?” [4].

As researchers in the linguistic field, we attempt to analyze the reasons behind this from a linguistic perspective. Research has found that the readability of a text can reflect its quality, which means that the higher the readability, the higher the quality of the text [5]. Additionally, readability is one of the factors that affect citation rate and is positively correlated with it [6, 7]. Moreover, texts with higher readability are less likely to be plagiarized or copied [8]. Furthermore, it has been found that articles with higher readability have a higher chance of being published in top-tier journals and of winning awards [9–10]. Therefore, the level of readability of a text reflects its quality to a certain extent, influences the spread of the text, and determines the degree to which it would be accepted.

Therefore, this study attempts to use the Coh-Metrix tool to measure the readability of texts written by Chinese and English native authors, to compare and analyze the readability-related parameters with Graesser & McNamara’s multilevel reading theory framework to explore whether there are significant differences in the readability of the two groups at different levels, and to analyze the possible reasons behind such differences [11]. Ultimately, this study aims to discover the linguistic factors that influence the citation rate of Chinese scholars’ SCI papers and to provide guidance for improving the quality of their international journal writing. The research expects to provide enlightenments to improve the citation rate and international influence of Chinese scientific research.

## II. LITERATURE REVIEW

Readability, also known as easibility, refers to the degree to which a text is easy to read and understand [12]. Graesser has identified 8 main components through principal component analysis to measure the readability of an article [13]. These main components are matched with different levels in Graesser and McNamara’s multilevel analyses of text characteristics: (1) the word level is associated with word concreteness; (2) the syntax level is associated with syntactic simplicity; (3) the textbase level is associated with referential cohesion; (4) the situation model level is associated with deep cohesion, verb cohesion, connectivity, and temporality; (5) the genre and rhetorical

Manuscript received April 5, 2023; revised May 16, 2023; accepted July 17, 2023.

The authors are with the School of Foreign Studies, Northwestern Polytechnical University, China.

\*Correspondence: uibeyuhang@163.com (H.Y.)

structure level corresponds to the narrativity of the article. Each main component also contains different variables [14] (see Table I).

TABLE I: READABILITY PARAMETERS AND POSITIVE AND NEGATIVE LOAD VARIABLES IN COH-METRIX

Positive Load Variables in Coh-Metrix		
	Readability parameters	Positive load variables
Word	Word concreteness	Concreteness for content words, mean; Imagability for content words, mean; Meaningfulness for content words, mean
Syntax	Syntactic simplicity	Causal verb incidence; Intentional verbs or adverbs incidence; LSA overlap, all sentences in paragraph, mean
Textbase	Referential cohesion	Content word overlap; Argument overlap; Noun overlap; Stem overlap; LSA;
	Depp cohesion	All connectives incidence; Causal connectives incidence; Logical connectives incidence; Temporal connectives incidence; Ratio of casual particles to causal verbs; Ratio of intentional particles to intentional verbs
Situation model	Verb cohesion	Polysemy for content words, mean; Verb overlap- adjacent sentences, mean LSA verb overlap
	Connectivity	Additive connectives incidence; Adversative and contrastive connectives incidence
	Temporality	Temporal cohesion, tense and aspect repetition, mean
Genre and rhetorical structure	Narrativity	Verb incidence; Adverb incidence; Pronoun incidence; CELEX Log frequency for all words, mean; CELEX Log minimum frequency for content words, mean; Familiarity for content words, mean; Negation density, incidence
Negative Load Variables in Coh-Metrix		
Word	Word concreteness	/
Syntax	Syntactic simplicity	Sentence length, number of words, mean; Left embeddedness, words before main verb, mean
Textbase	Referential cohesion	Type-token ratio; Lexical diversity for all words; Lexical diversity for verbs
	Depp cohesion	/
	Verb cohesion	/
	Connectivity	/
	Temporality	/
Genre and rhetorical structure	Narrativity	Noun incidence; Adjective incidence; Word length, number of syllables, mean; Age of acquisition for content words, mean; Agentless passive voice density, incidence; Noun phrase density, incidence

The values of each parameter in Table I are obtained by weighted summation of the corresponding variables. The negative load variables are transformed to ensure that the higher the numerical value of each variable, the higher the score of each parameter, and the higher the readability of the text.

Studies on academic readability mainly discuss the relationship between readability and academic influence, writing quality, and citation rate. For example, Oliver and Dallas *et al.* [5] used the Flesch measure to analyze the readability of articles and evaluated the quality of articles through manual scoring. Their results showed that the higher

the Flesch score, the higher the human score, which indicated that the higher the readability, the higher the quality of the article. Dowling and Hammami *et al.* [6] also measured the readability of journal article abstracts through

FOG and SMOG and analyzed their relationship with citation rate. Their research found that the readability of articles has a positive impact on their citation rate. Bryan [7] also pointed out that text readability is significantly related to citation rate. His data showed that an increase in readability reduced reading difficulty and thereby increased the citation rate of articles. Sun [8] analyzed the relationship between the score of paraphrasing, plagiarism, and other writing strategies and text readability. The results showed that authors are more likely to choose easy-to-read texts (i.e., text with higher average word frequency and shorter average sentence length) when paraphrasing or summarizing. This is because only when authors fully understand the content, can they process and apply them to their writing. For texts with low readability, authors may find it difficult to understand their content; therefore, may simply copy the original text, thereby increasing the probability of plagiarism. Sawyer *et al.* [9] found that the average word length and sentence length of award-winning papers were less than those papers that did not win awards, because the former are easier for readers to read and understand, and thus more readable. Fages [10] also found that the higher the readability of the text (i.e., the higher the FRES index and the lower the DCS, FKS, FOG, and SMOG indices, the lower the vocabulary and sentence difficulty), the higher the chance of being published in top journals. It can be seen that the readability of the text is not only one of the factors that measure the quality of academic writing, but also to some extent affects the dissemination and recognition of articles (such as citation rate, awards, and publication status) [15].

Given the positive relationship between readability and academic influence and citation rate, this study will analyze the similarities and differences of readability-related parameters between Introduction written by Chinese writers and English native writers at the linguistic level through self-built corpora of SCI journal article Introduction, attempting to discover the linguistic factors that affect citation rate and provide references for Chinese authors to improve the citation rate and academic influence of their articles.

### III. RESEARCH METHOD

#### A. Corpus Selection

In this study, 100 research papers in the fields of material science, control engineering, power energy, computer science, electronic information, management science, and mechanical manufacturing published in the first and second quarters English SCI journals from 2018 to 2022 were randomly selected, including 50 papers by Chinese authors and 50 papers by English native authors [16]. The standard for determining the author's nationalities was based on the author's affiliation provided in the article. For co-authored papers, the first author's nationality was taken. At the same time, only articles with the explicit word "Introduction" were selected, and the Introduction part was collected because the Introduction requires a solid language foundation and strong

logical expression ability. After the target articles were selected, two corpora were established, one for Chinese authors' Introduction (referred to as CIC) and the other for English native authors' Introduction (referred to as NIC). The total number of words in CIC was 52,102, and the total number of words in NIC was 45,356, which were comparable in size.

### B. Research Tool

The tools used in this study were mainly Coh-Metrix and AntConc. Coh-Metrix was developed by McNamara *et al.* [13] from the University of Memphis in the United States. It is a web-based text analysis tool that can automatically analyze the lexical, grammatical, and semantic features in the text. The parameters related to this study included referential cohesion, connectives, latent semantic analysis, lexical diversity, situation model, and text easibility principal component scores. Among them, the readability module included narrative, syntactic simplicity, lexical concreteness, referential cohesion, deep cohesion, verb cohesion, connectivity, and temporal aspects, which were analyzed based on Graesser's principal component analysis.

AntConc is a corpus retrieval tool developed by Laurence Anthony. In this study, its concordance and concordance plot functions were used to retrieve the location and contextual context of related terms, providing corresponding example sentences for the analysis of each parameter.

## IV. RESULTS AND DISCUSSION

Table II presents the Z score means, standard deviations, and independent sample t-test results of eight readability parameters for the Introduction of research articles written by Chinese authors and English native authors. According to Table II, the narrativity, syntactic simplicity, word concreteness, referential cohesion, and verb cohesion scores of Chinese authors' Introduction are higher than those of English native authors, while the deep cohesion, connectivity, and temporality scores are lower than those of English native writers' (see Table II for details). The differences between the two are significant in referential cohesion ( $p < 0.01$ ), deep cohesion ( $p < 0.05$ ), and temporality ( $p < 0.01$ ), but not significant in other parameters. According to the reading theory framework proposed by Graesser and McNamara, referential cohesion is a readability index at the textbase level, while deep cohesion and temporality are readability indices at the situation model level [11]. Therefore, the readability of Introductions written by Chinese and English native authors differ at different levels. The following section will provide a detailed discussion of these differences.

### A. Textbase-Referential Cohesion

Table III presents the positive load variable data of referential cohesion and the differences between Chinese and English native authors' Introductions. Regarding referential cohesion, Chinese authors' Introduction scored higher than English native authors' in all eight indicators, and there were significant differences between them (see Table III for details). Given that referential cohesion reflects readability at the textbase level, the data indicated that the readability of Chinese authors' Introduction is higher than that of English

native authors' at the textbase level.

TABLE II: READABILITY PARAMETERS OF INTRODUCTION OF CHINESE AND ENGLISH NATIVE WRITERS

	Chinese writers (mean±standard deviation)	English native writers (mean±standard deviation)	<i>p</i>
Narrativity Z score	-1.68±0.32	-1.69±0.37	0.882
Syntactic simplicity Z score	-0.20±0.47	-0.36±0.58	0.115
Word concreteness Z score	-0.40±0.64	-0.47±0.58	0.546
Referential cohesion Z score	0.04±0.73	-0.67±0.66	0.000**
Deep cohesion Z score	0.08±0.63	0.35±0.73	0.049*
Verb cohesion Z score	-1.12±0.68	-1.17±0.74	0.717
Connectivity Z score	-1.72±0.78	-1.57±0.88	0.365
Temporality Z score	-0.48±0.72	-0.11±0.68	0.009**

Halliday and Hasan divided cohesive devices into reference, substitution, ellipsis, conjunction, and lexical cohesion, among which reference includes personal reference, demonstrative reference, and comparative reference [17, 18]. Jiang Jinlin summarized the dimensions related to cohesion in Coh-Metrix, among which referential cohesion refers to the consideration of the cohesion of the text from the overlapping indicators of words, clauses, sentences, and paragraphs and latent semantic analysis. The higher the referential cohesion, the higher the readability of the text [14]. In second-language writing, authors usually engage in thinking activities in both their first language and second language and tend to use their first language in process control, structure organization, and content planning [19]. Furthermore, the linguistic form and discourse ability of their first language can also affect second-language writing [20]. It is thought that Chinese pays attention to functional meaning and semantic coherence and its cohesion is mainly achieved through semantic relations, such as the use of lexical overlap and ellipsis. On the contrary, English pays attention to structural form and grammatical cohesion, which needs various connective relationships and linguistic forms, such as reference and substitution (Examples (1) and (2)) [21, 22]. The data from this study show that the referential cohesion score of Chinese authors' Introductions is higher than that of English native authors', which may be due to the influence of Chinese way of thinking, leading to more repetition of nouns, arguments, content words, and stems in adjacent sentences and throughout the text (see Table III).

However, a text with good coherence does not necessarily require the use of a large number of cohesive devices, and frequent use of cohesive devices in the text does not necessarily ensure coherence [23]. Chinese authors, influenced by the transfer of their first language cohesive devices, tend to repeat excessively. Although this may increase the frequency of referential cohesion and readability scores, it does not necessarily mean an improvement in the quality of the article, as the excessive use of vocabulary repetition can lead to linguistic redundancy and cumbersome

expression in discourse [24]. In addition, studies have shown that referential cohesion is the most commonly used cohesive device in low-level English discourse Examples (3) and (4) [25].

TABLE III: POSITIVE LOAD VARIABLES OF REFERENTIAL COHESION IN INTRODUCTION OF CHINESE AND ENGLISH NATIVE WRITERS

	Nationality (mean±standard deviation)			p
		1.0 (n = 50)	2.0 (n = 50)	
Referential cohesion	Noun overlap- adjacent sentences	0.63±0.12	0.56±0.13	0.010**
	Argument overlap- adjacent sentences	0.69±0.11	0.63±0.12	0.018*
	Content word overlap- adjacent sentences	0.12±0.04	0.10±0.03	0.000**
	Argument overlap- all sentences	0.56±0.12	0.49±0.12	0.005**
	Stem overlap- all sentences	0.63±0.12	0.57±0.11	0.010**
	Content word overlap- all sentences	0.09±0.03	0.07±0.02	0.000**
	LSA overlap- all sentences in paragraph	0.08±0.02	0.08±0.02	0.596
	LSA overlap- all sentences	0.27±0.06	0.24±0.05	0.007**
	LSA overlap- adjacent sentences	0.29±0.06	0.26±0.05	0.011*

Example (1)

Active depth sensors (eg. Kinect V2) have been preferred over passive ones for a wide variety of use cases and have been studied extensively.

(NIC 31)

Example (2)

Liao and Axinte (2016a) reported that cutting with the worn and broken tool would result in secondary trauma of the bone for the avoidance of which they developed a force and acoustic emission-based process monitoring to detect the tool malfunction during the bone cutting.

(NIC 9)

Example (3)

Alternatively, the analog ULVPLLs, including the charge pump-based PLLs (CPPLL) and hybrid digital PLL (HDPLL), can avoid the design issues of the ULV TDC. However, they suffer from several other design challenges...

(CIC 19)

Example (4)

As a result, the iteration space at each feed position was enlarged, resulting in repeated iterations and invalid calculations in the iteration process.

(CIC 40)

The lexical cohesion index in Coh-Metrix is based on the lexical cohesion approach proposed by Halliday and Hasan [17], which employs lexical repetition, synonyms, antonyms, hyponyms, and hypernyms to achieve discourse coherence. Therefore, the lexical cohesion index here refers to semantic cohesion. Deerwester proposed Latent Semantic Analysis (LSA), which calculates the degree of semantic relatedness between linguistic units, including synonyms, antonyms, hyponyms, compound words, and semantic similarity between other words used in similar contexts [26]. LSA between sentences is a latent semantic cohesion index

that can reflect the semantic coherence of discourse and is an effective way to predict discourse readability [27]. The higher the LSA, the stronger the semantic coherence of the discourse, and the higher the reader’s reading speed and accuracy, that is, the higher the text readability [28].

The results of this study found that the LSA between sentences in the Introductions of Chinese authors is higher than that of authors from English native countries (see Table III). This indicates that the Introduction by Chinese authors has a higher potential semantic coherence, and the text is more readable. However, the semantic representation of discourse largely depends on the selection of words, and the requirements for syntax are not high [29]. When the lexical repetition rate and potential semantics in the discourse are both high, on the one hand, it indicates high textual coherence, but on the other hand, it also raises issues such as repetitive argumentation, similar content, and insufficient argumentation [30]. Therefore, the higher LSA of the Introduction of Chinese authors may also be caused by semantic singularity (i.e., the frequent repetition of the same words), but not a result of the improvement in readability or an improvement in the quality of the article.

TABLE IV: NEGATIVE LOAD VARIABLES OF REFERENTIAL COHESION IN INTRODUCTION OF CHINESE AND ENGLISH NATIVE WRITERS

	Nationality (mean±standard deviation)			p
		1.0 (n = 50)	2.0 (n = 50)	
Referential cohesion	Type-token ratio	0.40±0.07	0.45±0.06	0.000**
	Lexical diversity for all words	82.68±23.05	107.02±27.73	0.000**
	Lexical diversity for verbs	91.77±21.07	117.57±29.54	0.000**

Table IV summarizes the mean values and differences in negative load variables of referential cohesion between Introductions of Chinese writers and English native writers. The data show that Chinese writers’ Introductions scored lower than their English native counterparts in all three parameters, with extremely significant differences (p-values all <0.001, see Table IV), indicating lower lexical diversity in Chinese writers’ Introductions compared with English native writers’, and thus higher readability in terms of the textbase level.

Durán *et al.* have found a significant positive correlation between lexical diversity and reading difficulty [31]. In the present research, the type-token ratio of English native writers’ Introductions was 0.45, while that of Chinese writers’ Introduction was 0.40, indicating that English native authors repeated words less than Chinese writers and that the parameters for lexical and verb diversity were also higher for English native authors (Chinese and English native lexical diversity scores were 82.68 and 107.02, and verb diversity scores were 91.77 and 117.57 respectively, see Table IV). This suggests that English native writers’ Introductions have higher lexical diversity, lower repetition rates, greater comprehension difficulty, and lower readability. However, the fact seemed to be that lower readability did not affect their citation rate, and even some studies have shown that highly cited articles tend to have lower readability [32–36]. One reason may account for the higher readability of Chinese writers’ Introductions is the lower lexical diversity due to

their inclination to repeat words Examples (5) and (6), which is consistent with the results of most previous studies comparing lexical diversity between Chinese and English writing, and the main reason for Chinese authors' tendency to repeat words in English writing is their limited vocabulary [37, 38]. Therefore, although Chinese writers' Introductions may have higher readability based solely on referential cohesion, this does not necessarily reflect the true writing quality or higher citation rate.

Example (5)

This causes *very* low CP output impedance, and thus, makes the design of sub-0.3-V CP with an acceptable current matching performance be *very* challenging.

(CIC 19)

Example (6)

Although direct-RF polar TX has been invented and studied intensively for years, the direct-RF polar RX, which detects the phase and amplitude information of the received signal, as shown in Fig. 3, has been *barely* investigated with very *few* successful implementations.

(NIC 8)

B. Situation Model-Deep Cohesion, Temporality

Table V presents the mean values and differences in deep cohesion and temporality between Introductions of Chinese and English native writers. The data show that English native writers' Introductions had higher average values for intentional and temporal cohesion, and the differences between the two were significant ( $p < 0.01$ ). This indicates that Introductions of English naive writers had higher readability in terms of situational model.

TABLE V: DEEP COHESION AND TEMPORALITY PARAMETERS OF INTRODUCTION OF CHINESE AND ENGLISH NATIVE WRITERS

	Nationality (mean±standard deviation)		p
	1.0 (n = 50)	2.0 (n = 50)	
	Deep cohesion		
All connectives incidence	84.56±12.64	83.82±15.67	0.794
Causal connectives incidence	28.91±6.68	29.59±7.38	0.625
Logical connectives incidence	31.79±8.33	34.28±9.86	0.177
Temporal connectives incidence	13.93±5.12	13.82±5.19	0.912
Ratio of casual particles to causal verbs;	0.45±0.17	0.49±0.29	0.419
Ratio of intentional particles to intentional verbs	1.18±0.59	1.68±0.89	0.001**
Temporality			
Temporal cohesion	0.80±0.07	0.84±0.06	0.009**

Firstly, the significant differences in ratio of intentional particles to intentional verbs between the two indicate that, with a fixed number of intentional verbs, English native writers may use more intentional adverbs to enhance the intentionality of discourse [39], making the information contained in the discourse easier to understand and increasing readability. The significant difference in temporality shows that English native writers are better at unifying verb tense and verb aspect to express the chronological order of events in their Introductions, which helps readers better understand

the situation (see Examples (7) and (8). When temporal changes occur frequently without explicit words indicating the change in time (such as conjunctions, adverbs, noun phrases, etc.), the author's understanding of verb tense and verb aspect changes becomes more difficult. Since the Chinese language does not have the variation of verb tense and verb aspect, Chinese authors have lower mastery of it than English native writers, thereby reducing readability [11].

Secondly, English pays attention to a formal agreement, emphasizing the completeness of syntactic structure and often uses explicit logic connectors, such as conjunctions, to express semantic relationships in a context, whereas Chinese does not. Therefore, the use of causal conjunctions and logical conjunctions is higher in English native writers' Introductions than in Chinese authors' [23]. However, the so-called formal agreement or semantic agreement is based on frequency, and a formal agreement is not completely excluded in Chinese. As shown in Table V, the total number of conjunctions used in the Introductions of Chinese authors is even slightly higher than that of English native writers, taking temporal conjunctions as an example, Example (8). However, this may be due to the misuse and overuse of conjunctions by Chinese authors, which does not necessarily improve the coherence and cohesion of discourse [40, 41]. Therefore, the higher score of Chinese authors in a formal agreement may not indicate a higher readability literally.

Example (7)

Furthermore, the entirety of the optimization workflow *is to be carried out using* the capabilities of Galaxy Simulation Builder (GSB).

(NIC 3)

Example (8)

*Then* a most trustworthy subset can be selected and expanded into the support set.

(CIC 10)

Overall, the above data indicate significant differences between Chinese and English native writers in terms of referential cohesion, deep cohesion, and temporality in the Introductions. Furthermore, there are significant differences in the readability of the two groups of texts at the level of textbase and situation model. Firstly, the referential cohesion score of the Chinese author's Introductions is higher than that of the English native writer's, and the difference between the two is significant, indicating that the Chinese author's Introduction has higher readability at the textbase level. The higher repetition frequency of various types of words used by Chinese authors in sentence pairs reduces the difficulty of reading but also reflects the lower diversity of vocabulary used [42]. Secondly, the deep cohesion score and temporality score of the English native writers' Introductions are both higher than those of the Chinese authors', and the difference between the two is significant (see Tables II and V), indicating that the readability of the English native writers' Introductions is higher than that of the Chinese author's at the level of situational model. This indicates that when causal relationships, logical relationships, or chronological order appear in the discourse, English native writers are better at using causal conjunctions, logical conjunctions, intentional verbs, temporal conjunctions, and other means to help readers understand causal events and chronological order, thus

making the discourse content easier for readers to grasp [18].

## V. CONCLUSIONS AND ENLIGHTENMENT

This study conducted a comparative analysis of the readability of Introductions in SCI English journal articles authored by Chinese and English native authors, in an attempt to explore the reasons behind the lower citation rate of Chinese scholars' research articles from the linguistic perspective. The results revealed significant differences in reference cohesion, deep cohesion, and temporality between Chinese and English native authors' Introductions. At the discursual level, the Introduction written by Chinese authors had higher readability, while at the situational model level, the Introduction written by English native writers had higher readability. In terms of referential cohesion, Chinese authors tend to use repetitive vocabulary and have a higher potential semantic meaning in their discourse, leading to higher readability. But the fact that this does not seem to result in a higher citation rate indicate that Chinese authors should reduce lexical repetition, expand their vocabulary, utilize diverse forms of expression, and enhance their pragmatic abilities in academic writing. As for deep cohesion, Chinese authors tend to use various conjunctions frequently, but their use of causal and volitional conjunctions was inferior to that of English native writers, which made the internal logic of their articles harder to understand and reduced readability. Therefore, Chinese authors are advised to avoid overusing conjunctions, standardize their use of conjunctions, and pay attention to the use of function words such as causal and volitional conjunctions to better express the meaning of verbs in the discourse and make the content easier to understand. In terms of temporality, English native authors used verb tense and verb aspect more uniformly and accurately, resulting in higher readability of their articles. Therefore, Chinese authors are recommended to strengthen the use of verb tense and various participles to clarify the temporal sequence within the discourse and thus enhance the readability.

Additionally, some data in this study showed that Chinese authors used more cohesion devices than English native writers. However, what really correlated with writing quality is the correct use of cohesive devices, rather than their high frequency. Improper use of cohesive devices may even reduce the readability of articles, affecting their quality [43]. Moreover, the impact of vocabulary and syntax on text readability should not be ignored, especially for non-native English writers. Improving grammatical proficiency can also contribute to the enhancement of readability. Therefore, Chinese authors should strive to improve their language standardization, enhance their writing quality, and pursue a higher publication rate in international journals to disseminate their academic achievements.

Overall, although some aspects of Chinese authors' texts have higher readability scores than those of English native writers, it does not necessarily indicate the corresponding parameter indicators. Therefore, the impact of readability on citation rate cannot be generalized. Moreover, academic views on the relationship between readability and citation rate are not uniform. Further researches can explore other factors that affect citation rate, such as writing quality, cultural factors, and so on.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## AUTHOR CONTRIBUTIONS

Meiying Song designed and guided the research, offered the research materials, provided the framework, drafted parts of the manuscript and revised the manuscript; Qiwei Dai cleaned the materials and analyzed the data, drafted parts of the manuscript; Hang Yu offered the technical guidance for data processing and helped to modify data analysis; all three authors had approved of the final version.

## ACKNOWLEDGMENT

We are grateful to all the co-authors for the concerted efforts to accomplish this article. Our sincere gratitude should also go to the reviewers for their comments and to all staffs of the conference for their patience.

## REFERENCES

- [1] C. Huang and C. Zou, "Introduction and analysis of writing characteristics of SCI papers for medical students: Taking the top international medical journal the lancet as an example," *Medical Education Research and Practice*, vol. 28 & 141, no. 1, pp. 130–137, 2020.
- [2] G. Xiao and Z. Zhang, "The multi-dimensional features of English scientific paper titles: An investigation on the top 100 cited SCI research papers," *Chinese Journal of Scientific and Technical Periodicals*, vol. 27, no. 10, p. 6, 2016.
- [3] J. Chen, X. Yu, and Z. Shi, "A contrastive genre analysis of the introduction section in English papers of SCI—Indexed domestic and international biology journals," *Chinese Journal of Scientific and Technical Periodicals*, vol. 30, no. 2, p. 6, 2019.
- [4] C. Mu, "Research on English writing for research publication purposes since the 1980s: Looking backward and forward," *Foreign Language World*, vol. 1, pp. 56–64, 2015.
- [5] B. Oliver, M. J. Dallas, and M. Eckman, "Communication of empirical knowledge: An investigation of readability and quality of research in textiles and apparel," *Clothing and Textiles Research Journal*, vol. 16, no. 1, pp. 27–35, 1998.
- [6] M. Dowling, H. Hammami, and O. Zreik, "Easy to read, easy to cite?" *Economics Letters*, vol. 173, pp. 100–103, 2018.
- [7] B. C. McCannon, "Readability and research impact," *Economics Letters*, vol. 180, pp. 76–79, 2019.
- [8] Y. C. Sun, "Does text readability matter? A study of paraphrasing and plagiarism in English as a foreign language writing context," *Asia-Pacific Education Researcher*, vol. 21, no. 2, pp. 296–306, 2012.
- [9] A. Sawyer, J. Laran, and J. Xu, "The readability of marketing journals: Are award winning articles better written?" *Journal of Marketing*, vol. 72, no. 1, pp. 108–117, 2008.
- [10] D. M. Fages, "Write better, publish better," *Scientometrics*, vol. 122, pp. 1671–1681, 2020.
- [11] A. C. Graesser and D. S. McNamara, "Computational analysis of multilevel discourse comprehension," *Topics in Cognitive Science*, vol. 2, pp. 371–398, 2011.
- [12] S. Li, "A survey of studies on readability," *Journal of PLA University of Foreign Languages*, 2000.
- [13] A. C. Graesser, D. S. McNamara, and J. Kulikowich, "Coh-Matrix: Providing multilevel analyses of text characteristics," *Educational Researcher*, vol. 5, pp. 223–234, 2011.
- [14] J. Jiang, "The application of Coh-matrix in foreign language teaching and research," *Foreign Languages in China*, vol. 13, no. 5, pp. 58–65, 2016.
- [15] L. Zhang and J. Li, "Text readability and teaching of academic paper writing," *Modern Linguistics*, vol. 7, no. 4, p. 7, 2019.
- [16] B. Kachru, "Standards, codification and sociolinguistic realism: English language in the outer circle," in *English in the World: Teaching and Learning the Language and Literatures*, R. Quirk and H. Widowson, Eds. Cambridge: Cambridge University Press, 1985, pp. 11–36.
- [17] M. A. K. Halliday and R. Hasan, *Cohesion in English*, London: Longman, 1976.

- [18] Y. Liu, "Cohesive overtness, register and textual coherence," *Journal of Sichuan International Studies University*, vol. 6, 2002.
- [19] W. Wang and Q. Wen, "An investigation into L1 use in the L2 writing process of tertiary-level English learners in China," *Journal of PLA University of Foreign Languages*, vol. 25, no. 4, pp. 64–76, 2002.
- [20] A. Cumming, "Writing expertise and second language proficiency," *Language Learning*, vol. 39, 1989.
- [21] Y. Zuo, "Cohesive devices in Chinese and English compared," *Foreign Language Teaching and Research*, vol. 3, pp. 37–80, 1995.
- [22] Y. Wang and L. Zhang, "A case study of reference cohesion in English and Chinese Tex.," *Journal of PLA University of Foreign Languages*, vol. 29, no. 6, pp. 8–12, 2006.
- [23] D. S. McNamara, "The epistemic stance between the author and reader: A driving force in the cohesion of text and writing," *Discourse Studies*, vol. 15, no. 5, pp. 579–595, 2013.
- [24] J. Zhang, "Research on development of coherence in university English writing based on Coh-metrix," *The World & Chongqing*, vol. 3, pp. 66–72, 2016.
- [25] Y. Li, "Comparative study of the cohesive devices in high-and low-score English argumentative essays from the perspective of cohesion theory," *English Teachers*, vol. 22, no. 5, pp. 47–51, 2022.
- [26] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harskman, "Indexing by latent semantic analysis," *Journal of the American Society for Information Science*, vol. 41, pp. 397–407, 1990.
- [27] W. Kintsch and D. Vipond, *Perspectives in Memory Research*, The MIT Press, 2020.
- [28] P. W. Foltz, W. Kintsch, and T. K. Landauer, "The measurement of textual coherence with latent semantic analysis," *Discourse Processes*, vol. 25, pp. 285–307, 1998.
- [29] S. Gui, "The theory of latent semantic analysis and its application," *Modern Foreign Languages*, vol. 26, no. 1, pp. 76–84, 2003.
- [30] L. He and Y. Sun, "Investigation of the effects of prompt characteristics on Chinese test-takers' integrated writing performance," *Foreign Language Teaching and Research*, vol. 2, pp. 237–250, 2015.
- [31] P. Durán, D. Malvern, B. J. Richards, and N. Chipere, "Developmental trends in lexical diversity," *Applied Linguistics*, vol. 25, no. 2, pp. 220–242, 2004.
- [32] J. Hartley, E. Sotto, and J. Pennebaker, "Style and substance in psychology: Are influential articles more readable than less influential ones," *Social Studies of Science*, vol. 32, no. 2, pp. 321–34, 2002.
- [33] S. Strememeh, I. Vemiers, and P. C. Verhoef, "The quest for citations: Drivers of article impact," *Journal of Marketing*, vol. 71, no. 3, pp. 171–193, 2007.
- [34] A. Gazni, "Are the abstracts of high impact articles more readable? Investigating the evidence from top research institutions in the world," *Journal of Information Science*, vol. 37, no. 3, pp. 273–281, 2011.
- [35] F. Didegah and M. Thelwall, "Which factors help authors produce the highest impact research? Collaboration, journal and document properties," *Journal of Informetrics*, vol. 7, pp. 861–873, 2013.
- [36] S. Dolnicar and A. Chapple, "The readability of articles in tourism journals," *Annals of Tourism Research*, vol. 52, pp. 161–166, 2015.
- [37] S. H. Kwon, "Lexical richness in L2 writing: How much vocabulary do L2 learners need to use?" *English Teaching*, vol. 64, no. 3, pp. 155–174, 2009.
- [38] X. Pan and F. Yuan, "Quantitative comparison of the language features in abstracts of academic papers written by Chinese and American scholars of science and engineering," *Journal of Hainan Tropical Ocean University*, vol. 28, no. 1, pp. 114–120, 2021.
- [39] W. Jia and P. Zhang, "Exploring the relationship between textual features and writing quality in TEM8 writing test," *Foreign Language Testing and Teaching*, no. 1, pp. 1–9, 2020.
- [40] W. J. Crewe, "The illogic of logical connectives," *ELT Journal*, vol. 44, pp. 316–325, 1990.
- [41] J. Milton and S. C. Tsang, "A corpus-based study of logical connectors in EFL students' writing: Directions for future research," *The Hong Kong University of Science and Technology*, 1993.
- [42] M. Linnarud, *Lexis in Composition: A Performance Analysis of Swedish L2 learner's Written English*, Malmo: CWK Gleerup, 1986.
- [43] M. Heilman, K. Collins-Thompson, C. Jamie *et al.*, "Combining lexical and grammatical features to improve readability measures for first and second language texts," in *Proc. Human Language Technology Conference of the North American Chapter for the Association of Computational Linguistics*, Rochester, New York, USA, April 2007, pp. 22–27.

Copyright © 2023 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).