

| RESEARCH ARTICLE**Multivocality and Readability in High School English Writing: The Role of Task Type****Haiyue Wu***College of Education, Zhejiang University of Technology, Zhejiang, China***Corresponding Author:** Haiyue Wu, **E-mail:** why1329133419@163.com**| ABSTRACT**

In certain genres of high school English writing, incorporating third-party voices can enrich the content of an essay, while also posing challenges to readability. Drawing on a small learner corpus, this study analyzed writing samples produced by 55 high school English learners across two task types: practical writing and continuation writing. Measures of multivocality and readability were calculated for each essay to examine whether and how the inclusion of third-party voices relates to text readability under different task conditions. The findings are as follows: (1) Overall, continuation writing involved richer multivocal dialogue and exhibited lower reading difficulty. Practical writing showed higher frequencies of proclaim and entertain resources, whereas continuation writing demonstrated higher frequencies of disclaim and attribute resources. (2) As multivocal resources were gradually added, the readability of practical writing followed a pattern of "decrease – increase – decrease." In contrast, the readability of continuation writing was not affected by multivocality. Practical writing that meets general readability standards tends to avoid using either too few or too many multivocal resources. This study provides empirical evidence on how high school students can use multivocal resources to produce clearer and more readable English essays.

| KEYWORDS

English writing tasks; multivocality; readability; correlation study

| ARTICLE INFORMATION**ACCEPTED:** 01 February 2026**PUBLISHED:** 09 February 2026**DOI:** 10.32996/jeltal.2026.8.3.2**1. Introduction**

Writing is a crucial component of language output. Currently, the unified enrollment examinations in China's basic education stage are placing increasingly higher demands on the writing proficiency of English learners. These assessments not only test students' ability in word and sentence formation but also comprehensively evaluate their higher-order thinking skills, such as logical and critical thinking (Lu, 2019). However, many English learners nowadays blindly pursue linguistic complexity, neglecting how to produce a well-structured and easily readable composition.

"Voice" is one of the fundamental components of excellent writing (Stewart, 1992). Through the use of voice, authors articulate their stance in multi-directional interactions with readers and third-party sources to achieve communicative purposes. Although the evaluation of voice quality in writing has attracted the attention of domestic scholars (Zhang, 2023), how it influences micro-level textual features such as readability remains underexplored. This disconnect makes it difficult for teaching practices to quantitatively assess the actual impact of multivocality on reading effectiveness. This study utilizes a corpus of writing samples from 55 high school English learners, focusing on the two most common task types they encounter—practical writing and continuation writing. By distinguishing between these two task types, the study calculates multivocality and readability metrics for each composition. It aims to examine whether the introduction of third-party voices by students leads to changes in text readability across different task types.

2. Literature Review

2.1 Research on Multivocality in Writing

Bakhtin's work laid the theoretical foundation for the study of voice in discourse, conceptualizing voice as a constellation of socially situated viewpoints and meanings realized through language (Bakhtin, 1984). Since then, the notion of voice has attracted sustained scholarly attention, particularly in literary studies, where texts are commonly understood as sites of interaction among multiple voices rather than as monologic expressions of a single authorial consciousness (Zhuo, 2015; Lai & Su, 2022). These voices engage with one another around shared events and themes, giving rise to what Bakhtin famously described as polyphonic dialogue.

Subsequent research has extended the concept of voice beyond literary discourse to written communication more broadly. In writing studies, voice has been closely associated with authorial stance, as writers position themselves in relation to readers and alternative viewpoints (Zeng et al., 2024). The presence of multiple voices thus expands the range of stance-taking resources available to writers. This line of inquiry has been particularly prominent in research on academic writing, a genre characterized by heightened dialogic complexity. Early studies frequently examined how sociocultural background, disciplinary conventions, and language proficiency shape the construction of voice in English learners' texts (Hirvela & Belcher, 2001; Zhang & Zhan, 2020). However, findings across these studies have remained inconclusive, partly due to differences in analytical frameworks and research focus.

In recent years, the Appraisal framework within Systemic Functional Linguistics has provided a more systematic approach to the study of multivocality in writing. Specifically, the Engagement system offers a principled account of how writers linguistically manage dialogic space through lexico-grammatical choices. Research adopting this framework has demonstrated that different engagement resources serve distinct dialogic functions. For example, Gao et al. (2022) found that highly proficient native speakers in hard sciences tended to favor expansive reporting verbs, thereby opening space for dialogic negotiation. Building on this perspective, scholars have proposed analytic frameworks for examining multivocality in academic discourse across languages, taking into account genre-specific and cultural characteristics (Zhang, 2023; Yu & Zhang, 2021). Despite these advances, empirical research on multivocality in high school L2 writing remains limited. Existing studies are largely theoretical in nature and often treat multivocality as a single, isolated variable, leaving its interaction with other dimensions of writing performance underexplored (Zhao, 2016).

2.2 Research on Factors Influencing Readability

Readability, also referred to as ease of reading, denotes the degree or quality of how easily a text can be read and understood (Li, 2000). It is a key indicator of information transmission efficiency. Texts with fluent expression and appropriately challenging reading materials make it easier for readers to comprehend the author's knowledge framework construction process, content, and the social value of research findings (Wang et al., 2020). In other words, texts with high readability have higher information transmission efficiency. Therefore, analyzing the factors influencing text readability can reduce cognitive load, minimize comprehension deviation, and thereby enhance the rate of information dissemination.

Linguistic features are core factors affecting text readability. Cheng et al. (2020) analyzed key factors influencing reading difficulty across four levels: character, word, sentence, and discourse, and constructed a text readability formula using linear regression equations. Among these, Chinese characters are the smallest units of the writing system, and their difficulty impacts text reading difficulty. Generally, a higher number of strokes and lower character frequency indicate a more uncommon character, making it more difficult to recognize. Words are the smallest meaningful units that can be used independently in a sentence. Low-frequency words and specialized terminology significantly increase cognitive load (DuBay, 2004). Part of speech directly influences sentence complexity; heavy use of adverbs can confuse second language learners (Hernández, 2006), and the correct use of function words is also a challenge for learners of Chinese as a second language (Li, 1998). At the syntactic level, long sentences and nested structures reduce readers' text processing efficiency. For instance, an overuse of passive voice can increase difficulty indices by up to 30% (McLaughlin, 1969). Graesser et al. (2015) argue that text cohesion and coherence significantly influence readability. They suggest that when analyzing text readability, deep discursive semantic features should be incorporated alongside lexical and syntactic characteristics. Wu et al. (2020) categorized potential discourse-level factors affecting Chinese text readability into text complexity and text cohesion, incorporating them into an indicator system and comparing their predictive power for texts of different levels and difficulty grades. Readability is essentially the compatibility between linguistic form and readers' cognitive abilities. Therefore, it is necessary to study the factors influencing text readability to better match texts to readers and facilitate information dissemination.

3. Research Design

3.1 Research Questions

This study aims to answer the following two questions:

- (1) What differences exist in the use of multivocal resources and readability in the English writing texts of high school learners across different task types?
- (2) Does the readability of high school English learners' writing texts change with the introduction of multivocal resources under different task types?

3.2 Data Collection

The research corpus consists of 110 essays written by 55 high school English learners from a high school in Suzhou. All participants were taught by the same instructor. The researcher collected one practical writing sample and one continuation writing sample from each participant during two separate weekly tests. The two types of writing tasks were administered with a two-week interval, and both required approximately 150 words.

Regarding task type, the continuation writing task not only requires students to depict events occurring in different scenarios but also comprehensively tests their reading comprehension and inferential abilities, thus involving a complex task. According to Robinson's (2001) Cognition Hypothesis, its task complexity along resource-directing dimensions is higher than that of the letter of advice, which primarily assesses writing ability.

3. Data Measurement

To address the first research question, this study annotated and analyzed each essay based on the Heteroglossia System within the Engagement framework of Appraisal Theory (Martin & Rose, 2003), which provides a principled account of how writers manage dialogic space. In this study, heteroglossia is operationalized as textual multivocality, referring to the extent to which writers introduce and negotiate alternative voices in the construction of meaning. Drawing on Bakhtin's dialogic principle, such multivocality is realized through linguistic resources including projection, modality, and concession (Zhang, 2024). Based on the scope of dialogic negotiation, heteroglossia are primarily categorized into four types: Contraction: Disclaim, Contraction: Proclaim, Expansion: Entertain, and Expansion: Attribute. Figure 1 illustrates these four types of multivocal resources along with their typical expressions. For statistical purposes, instances of contractive heteroglossia were coded as 1 point, while expansive heteroglossia was coded as 2 points. The scores for each essay were then summed to yield an overall multivocality score, representing the relative size of the dialogic space constructed in the composition.

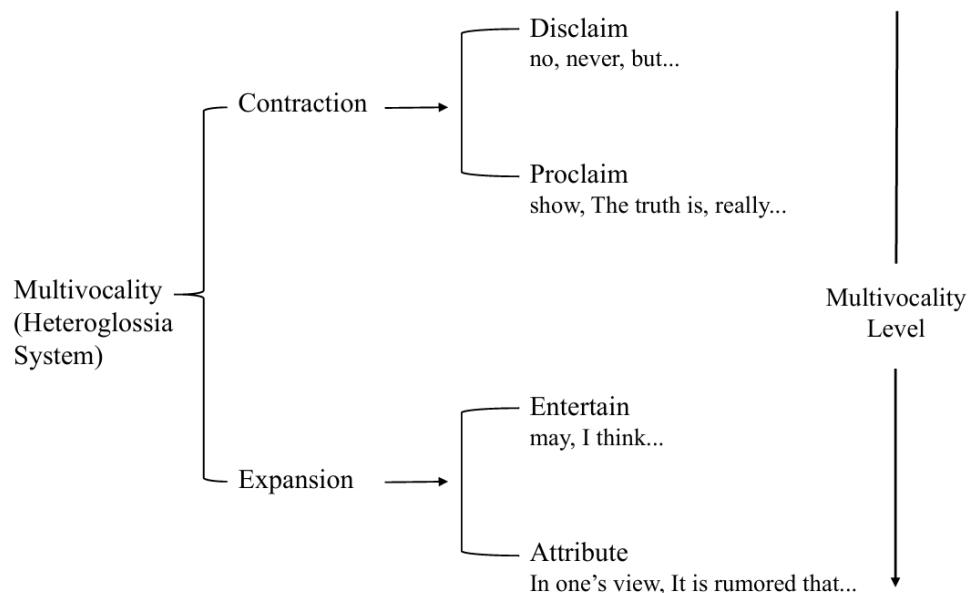


Figure 1. Framework of the Multivocality (Heteroglossia System)

For the second research question regarding readability, two widely used computational formulas will be adopted: the Flesch-Kincaid Index and the FOG Index. Higher scores in the Flesch-Kincaid Index and FOG Index indicate lower readability. The researchers will input the essay data into R software and appropriately adapt the analysis code provided by Shi Yaqian and Lei (2023) to calculate the readability metrics. The specific calculation formulas are detailed in Table 1. Finally, to examine the relationship between multivocality and readability, regression models will be constructed in R using the lm() function to analyze the associations between multivocality and the Flesch-Kincaid Index, as well as between multivocality and the FOG Index.

Table 1 Calculation Formulas for Two Readability Metrics

	Formula	Description
Flesch-Kincaid	$(0.39 \times \text{Average Sentence Length}) + (11.8 \times \text{Average Syllables per Word}) - 15.59$	higher value indicates higher difficulty
FOG	$0.4 \times (\text{Average Sentence Length} + \text{Percentage of Complex Words})$	higher value indicates higher difficulty

4. Results

4.1 Multivocal Resources and Readability in Writing Texts Across Different Task Types

As shown in Table 2, a total of 565 multivocal resources were identified across the 110 essays. A comparison reveals that in both writing task types, expansive multivocality (63.3% / 53.8%) was more frequent than contractive multivocality (36.7% / 46.2%). In practical writing, entertain and proclaim resources accounted for a significant proportion, representing 38.4% and 35.5% of the total multivocal resources in this task type, respectively. Disclaim resources were notably scarce, constituting only 1.2% of the total. Conversely, in continuation writing, attribute and proclaim resources comprised the largest shares, making up 37.2% and 23.4% of the multivocal resources in this task type, respectively. Entertain resources were relatively less common, accounting for 16.6%. This indicates that the practical writing (letter of advice) expands the dialogic space through personal viewpoints and inferences on one hand, while on the other hand, it employs endorsing and concurring vocabulary to reinforce the author's stance. In contrast, continuation writing opens the dialogic space by depicting character communication and debate, while simultaneously negating or supplementing adjacent propositions, thereby guiding readers towards the central theme through critical engagement.

Table 2 Statistics of Multivocal Resources Across Different Task Types

		Practical Writing (Low Complexity)	Continuation Writing (High Complexity)	Example
Contraction	Disclaim	3 (1.2%)	75 (23.4%)	No one could give him an answer.
	Proclaim	87 (35.5%)	73 (22.8%)	Here's a truth that more people prefer to learn Chinese.
	Subtotal	90 (36.7%)	148 (46.3%)	
Expansion	Entertain	94 (38.4%)	53 (16.6%)	From my perspective, I also suggest you do so.
	Attribute	61 (24.9%)	119 (37.2%)	Ella's words echoed in his mind, letting Mark to make a decision.
Total	Subtotal	155 (63.3%)	172 (53.8%)	
		245 (100%)	320 (100%)	

As shown in Table 3, the high-complexity task (continuation writing, $M=11.69$) exhibits significantly more multivocality than the low-complexity task (practical writing, $M=7.27$). This aligns with the total counts in Table 2 (245 / 320), indicating that students introduce third-party voices more frequently in the more complex continuation writing task. According to both the Flesch-Kincaid Index and the FOG Index, the mean scores for the high-complexity task (continuation writing) are lower than those for the low-complexity task (practical writing), suggesting that continuation writing has higher readability. Generally, readability scores around 7–8 indicate that a text is suitable for a general audience. Thus, based on the two readability metrics in Table 3, it can be observed that the practical writing produced by students is generally appropriate for public readership, while the continuation writing appears relatively simpler for its intended audience.

4.2 The Relationship Between Multivocality and Readability in Writing Texts Across Different Task Types

In practical writing, a moderate negative correlation exists between multivocality and both the Flesch-Kincaid Index ($P=0.007$, $r=-0.358$) and the FOG Index ($P=0.009$, $r=-0.348$). This indicates that in practical writing, the presence of other voices tends to reduce these two indices to some extent. Since lower Flesch-Kincaid and FOG values signify higher readability, maintaining a certain dialogic space may help decrease the reading difficulty of practical writing. However, the moderate strength of these correlations also suggests the potential existence of non-linear relationships that have not been fully captured.

Table 3 Descriptive Statistics of Multivocality and Two Readability Metrics for Essays Across Different Task Types

Type		Mean	Sd	Min	Max	1	2	3
Practical Writing (Low Complexity)	Multivocality	7.27	2.39	2.00	16.00	-		
	Flesch-Kincaid	6.50	1.10	3.82	9.10	-0.358**	-	
	FOG	8.37	1.22	5.75	11.68	-0.348**	0.890**	-
Continuation Writing (High Complexity)	Multivocality	11.69	3.56	2.00	17.00	-		
	Flesch-Kincaid	4.15	1.11	1.43	6.91	-0.048	-	
	FOG	6.38	1.23	3.06	9.05	0.012	0.942**	-

As shown in Table 4, the study first fitted linear, quadratic polynomial, and cubic polynomial regression models using multivocality in practical writing as the independent variable and the Flesch-Kincaid Index as the dependent variable. Among these, the cubic polynomial model was the optimal one, explaining approximately 20% of the variance. Using the same approach, three regression models were fitted with multivocality as the independent variable and the FOG Index as the dependent variable, revealing that the cubic polynomial still explained the most variance. This indicates that the relationship between multivocality intensity and readability in practical writing generally follows a cubic curve (see Figure 2). In the initial stage of opening the dialogic space, multivocality intensity is positively correlated with readability: as occasional individual voices appear in the essay, readability indices increase, making the text harder to read. When the dialogic space crosses the first critical point and enters the second stage, multivocality intensity becomes negatively correlated with readability: as more voices are introduced, readability indices decrease, making the essay easier to read. When the dialogic space crosses the second critical point and reaches the third stage, multivocality intensity and readability again show a positive correlation: as voices further accumulate, readability indices increase, and the text's reading difficulty rises once more. When the multivocality score is 5 points, practical writing is the most difficult to read; when the score is 13 points, it is the easiest to read. Based on the two curves, it can be roughly estimated that multivocality scores between 4 and 10 points meet the readability requirements for a general audience.

Table 4 Summary of Curvilinear Regression Models for Multivocality and Readability Metrics in Practical Writing

Equation		R2	Adj.R2	sig.	Constant	b1	b2	b3
Flesch-Kincaid	Linear	0.11	0.09	0.01	7.612	-0.153		
	Quadratic	0.12	0.08	0.04	7.069	-0.011	-0.008	
	Cubic	0.20	0.15	0.01	3.390	1.549	-0.205	0.007
FOG	Linear	0.11	0.10	0.01	9.618	-0.171		
	Quadratic	0.11	0.08	0.04	9.312	-0.092	-0.005	
	Cubic	0.15	0.10	0.04	6.775	0.984	-0.140	0.005

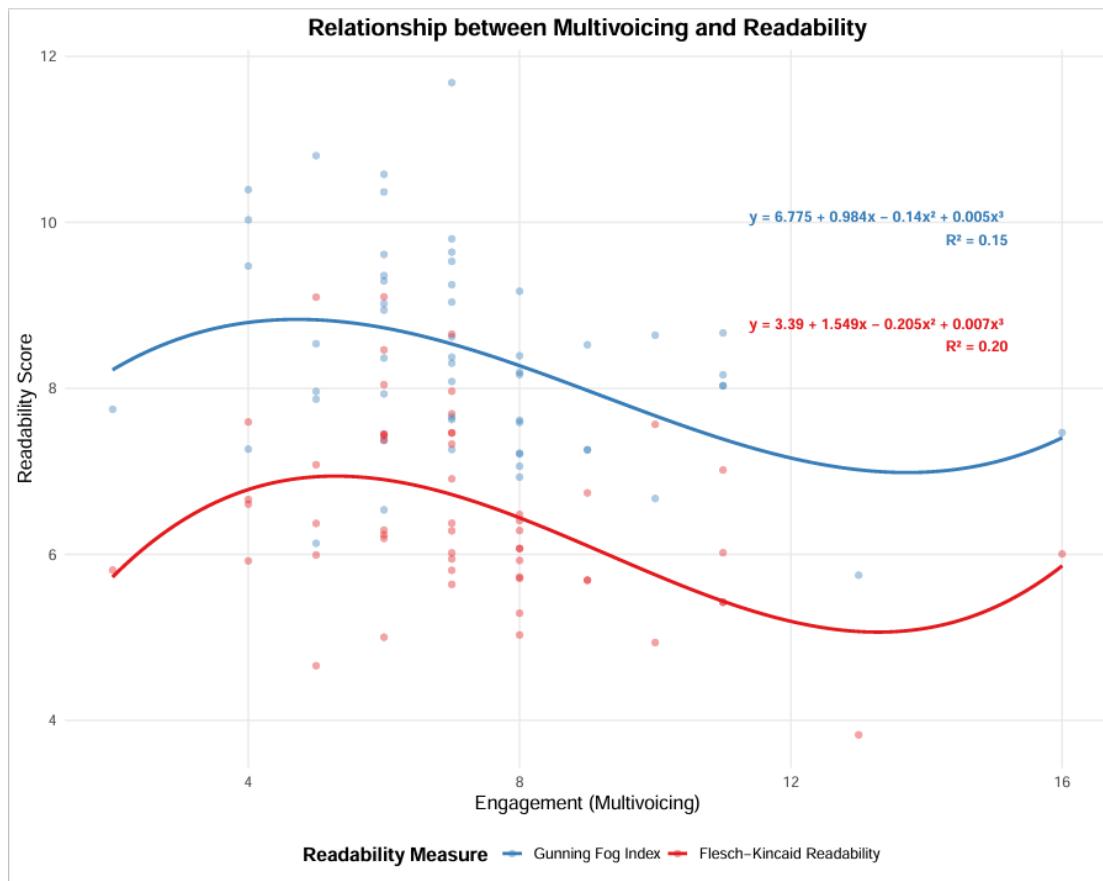


Figure 2. Fitted Curves of the Curvilinear Regression Models for Multivocality and Readability Metrics in Practical Writing

In continuation writing, no significant correlation is observed between multivocality and the Flesch-Kincaid Index ($P=0.727$, $r=-0.048$) or the FOG Index ($P=0.932$, $r=0.012$). Linear, quadratic polynomial, and cubic polynomial regression models were fitted using multivocality in continuation writing as the independent variable and the Flesch-Kincaid Index and FOG Index as dependent variables, respectively. The significance levels (P -values) for all resulting models were far greater than 0.05, indicating that in continuation writing, there is no statistically observable regression relationship between third-party voices and text reading difficulty. This finding stands in stark contrast to the results obtained for practical writing.

5. Discussion and Conclusion

This study investigated the differences in textual multivocality and readability under varying task complexities, as well as the influence of multivocality on readability, based on data from practical writing and continuation writing texts produced by 55 high school English learners. The main findings are as follows:

First, practical writing exhibited a higher frequency of proclaim and entertain resources, whereas continuation writing showed a greater use of disclaim and attribute resources. Overall, continuation writing involved a higher degree of multivocal dialogue. This distribution can be interpreted in light of the distinct communicative purposes underlying the two task types. Letters of advice aim to offer workable solutions and demonstrate their validity. Accordingly, students tend to foreground endorsed voices and personal viewpoints in order to make their suggestions more focused and persuasive. Excessive reliance on attribution without a clearly articulated stance, by contrast, may obscure authorial responsibility and weaken the intended communicative effect. Continuation writing, however, is oriented toward the development of coherent narratives with logically unfolding plotlines. Dialogue plays a central role in introducing conflict and advancing the storyline (Bai Qinghua, 2018). Students therefore make extensive use of multivocal interaction among characters to reveal information and propel the narrative forward. The relatively high frequency of disclaim resources in continuation writing can thus be understood as a linguistic reflection of textual conflict. In addition, continuation writing was found to exhibit higher readability than practical writing. This finding lends partial support to the Cognition Hypothesis, which suggests that increased cognitive demands along resource-directing dimensions may lead to reduced linguistic complexity in written output.

In low-complexity practical writing tasks, students frequently rely on template-based strategies. Familiarity with commonly used functional expressions enables them to meet task requirements efficiently, but it may also encourage the use of more complex vocabulary, thereby increasing reading difficulty (Zhang Yujie & Jiang Jingyang, 2020). By contrast, continuation writing requires on-the-spot text production following comprehension of the source material, leaving limited room for the direct application of memorized templates. Although such tasks involve frequent event description, students' lexical choices are largely confined to basic action verbs. Under time constraints, learners tend to prioritize readily accessible vocabulary, which may contribute to the comparatively lower reading difficulty observed in continuation writing.

Second, as multivocality increased, the readability of practical writing followed a non-linear pattern characterized by an initial decrease, a subsequent increase, and a final decrease. This suggests that, for texts intended for a general readership, multivocal resources are most effective when employed at a moderate to moderately high level. Both overly restricted and excessively expansive dialogic spaces may heighten readers' comprehension burden, whether through underdeveloped reasoning or fragmented logical progression. When the dialogic space falls within an optimal range, however, writers appear more capable of organizing viewpoints into a coherent argumentative structure. The strategic use of multivocal resources in such cases facilitates logical connectivity and helps readers follow the line of reasoning with greater ease.

In contrast, no observable relationship was found between multivocality and readability in continuation writing. Given the generally limited linguistic repertoire of high school English learners, continuation writing tends to rely on everyday vocabulary and basic verbal constructions. Multivocal interaction in narrative development is therefore achieved with minimal use of abstract logical connectors or generalized expressions. As a result, the incremental introduction of voices imposes little additional cognitive load on readers, leaving overall readability largely unaffected.

From a pedagogical perspective, these findings suggest that differentiated instructional approaches are needed for different writing task types. In practical writing, teachers may guide students toward a balanced use of multivocal resources, avoiding both excessive self-assertion and overreliance on external voices. In continuation writing, instruction may instead focus on encouraging more varied and precise verb use to enrich narrative expression.

This study contributes to a growing body of research on multivocality in second language writing by demonstrating how the deployment of voices interacts with readability under different task conditions. Nevertheless, several limitations should be acknowledged. The data were drawn from a single high school in one region, which may limit the generalizability of the findings. Future research could extend the scope to learners from diverse educational contexts and proficiency levels. Moreover, individual learner variables such as language proficiency and writing motivation were not systematically examined. Further studies may address these factors to provide a more nuanced account of the mechanisms through which multivocality influences readability in L2 writing.

Funding: The study reported in this paper represents a contribution to 'a Corpus- based Study on the Dialogism in High School English Classroom Discourse' (Grant No.: SJCX25_1810) funded by Jiangsu Provincial Department of Education.

References

- [1] Bakhtin, M. M. (1984). *Problems of Dostoevsky's poetics*. University of Minnesota Press.
- [2] Bai, Q. (2018). The reference of Bakhtin's polyphonic novel theory to China's drama creation. *Gansu Social Sciences*, (02), 108–112.
- [3] Cheng, Y., Xu, D., & Dong, J. (2020). Analysis of key factors in text reading difficulty grading and research on readability formulas based on Chinese textbook corpus. *Applied Linguistics*, (01), 132–143. (in Chinese)
- [4] DuBay, W. H. (2004). *The principles of readability*. Impact Information.
- [5] Gao, J., Zhang, M., & Chen, W. (2022). A revisit to the development of reporting verb use driven by engagement theory: An interdisciplinary/level study based on a corpus of high-level native learners. *Foreign Languages Research*, 39(05), 52–56+82.
- [6] Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2015). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234.
- [7] Hernández, M. S. (2006). The position of adverbs in English: Trying to solve a major problem most language learners usually face. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 32(1), 271–285.
- [8] Hirvela, A., & Belcher, D. (2001). Coming back to voice: The multiple voices and identities of mature multilingual writers. *Journal of Second Language Writing*, 10(2), 83–106.
- [9] Lai, L., & Su, S. (2022). Dialogism in literary history discourse from the perspective of engagement theory. *Journal of Shanghai Jiaotong University (Philosophy and Social Sciences)*, 30(02), 127–136.
- [10] Li, S. (2000). An survey of studies on readability. *Journal of PLA University of Foreign Languages*, 23(4), 1–5.
- [11] Li, X. (1998). On the teaching of function words in teaching Chinese as a foreign language. *Chinese Teaching in the World*, (03), 65–71.

- [12] Lü, G. (2019). Analysis of English writing test in senior secondary school entrance examination and revelation found. *Teaching and Administration*, (28), 45–48.
- [13] McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading*, 12(8), 639–646.
- [14] Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, 22(1), 27–57.
- [15] Shi, Y., & Lei, L. (2023). *Linguistic digital humanities and R language practice*. Shanghai Jiao Tong University Press.
- [16] Stewart, D. C. (1992). Cognitive psychologists, social constructionists, and three nineteenth-century advocates of authentic voice. *Journal of Advanced Composition*, (2), 279–290.
- [17] Wang, N., & Luo, J. (2020). Non-academic impact of scientific research and its assessment: What, why and how? *Journal of East China Normal University (Educational Sciences)*, 38(4), 62–71. (in Chinese)
- [18] Wu, S., Yu, D., & Jiang, X. (2020). Development of linguistic features system for Chinese text readability assessment and its validity verification. *Chinese Teaching in the World*, 34(01), 81–97.
- [19] Yu, H., & Zhang, S. (2021). A study on heteroglossia in Chinese academic discourse. *Contemporary Rhetoric*, (06), 49–59.
- [20] Zeng, X., Liu, J., & Bowen, N. E. J. A. (2024). A comparative study of voice in Chinese English-major undergraduates' timed and untimed argument writing. *Assessing Writing*, (62), 100896.
- [21] Zhang, D. (2023). Voice in academic research articles: Shift of conceptual focus and the establishment of a tripartite analytical framework. *Foreign Language and Literature*, 40(03), 98–110+136. (in Chinese)
- [22] Zhang, D. (2024). New trends in the study of engagement resources in academic articles. *Modern Foreign Languages*, 47(06), 849–858.
- [23] Zhang, F., & Zhan, J. (2020). Understanding voice in Chinese students' English writing. *Journal of English for Academic Purposes*, (45), 100844.
- [24] Zhang, Y., & Jiang, J. (2020). Impact of task complexity on complexity and accuracy of L2 written production. *Journal of Xi'an International Studies University*, 28(04), 49–54.
- [25] Zhao, C. G. (2017). Voice in timed L2 argumentative essay writing. *Assessing Writing*, (31), 73–83.
- [26] Zhuo, M. (2015). Elegy in monologue: Monologic single-voiced discourse in *The Sheep Set Free*. *Contemporary Writers Review*, (04), 62–68.