

An Evaluation of the Assessment Measure for Novice L2 Learners' English Writing

Wang Yumin*, Jia Lijuan, Zhao Jingxuan

International Education College, Henan University, China

Received April 22, 2019; Revised November 11, 2019; Accepted November 19, 2019

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Academic writing poses a consistent challenge in many students' university career. This study conducted a systematic diagnostic assessment to identify specific discourse strengths and weaknesses in first-year undergraduate students' English writing. Our study found that the top three prevalent weaknesses are related to overall discourse awareness, syntactic proficiency, and idea development. Based on the knowledge of students' specific strengths and weaknesses, more targeted remediation prescription can be designed and delivered for maximal support in facilitating first year undergraduate students' development of English writing proficiency in the classroom setting.

Keywords Diagnostic Assessment, Chinese EFL Learners, Discourse Features

1. Introduction

Academic writing in a second or foreign language (L2) frequently poses a challenge in many people's academic lives. In the tertiary level education in China, many students have accumulated certain knowledge of vocabulary and grammar of English, but results in their performance and proficiency tests (as seen in statistics released in the IELTS website) suggest that their writing in English seems somewhat stagnated despite the persistent efforts from both students and their teachers. Diagnostic assessment, i.e., the act of precisely analyzing a problem and identifying its causes for the purpose of effective treatment (Rupp, Templin & Henson, 2010), arises as a helpful approach to cope with this situation.

Diagnostic language assessment is designed to identify the strengths and weaknesses in individuals' or group's knowledge and use of language. Alderson (2005) specified that "focusing on strength will enable the identification of the level a learner has reached, and focusing on weaknesses

or possible areas for improvement should lead to remediation or further instruction" (p.257). What makes diagnostic assessment stand out from other types of assessment is that diagnostic assessment not only identifies problems, but also and more importantly, searches for underlying causes of the problems, especially the persistent or recurring problems so that appropriate remediation activities can be prescribed (McMillan, 2014; Rupp, Templin, & Henson, 2010).

Given Chinese students' persistent problems and difficulties in English writing, the present study aims at developing a set of fine-grained writing rubrics and further applying it to help diagnose the specific discourse strengths and weaknesses as demonstrated in first year undergraduate students' English writing. In the following part, before reporting our findings, we first review existing literature to gain a systematic conceptualization of discourse features as conceived in related assessment rubrics and synthesize latest findings on Chinese undergraduate students' discourse problems in their writing.

2. Literature Review

2.1. Conceptualization of Discourse Features in Writing

"Discourse" has been a buzz word in many social science studies: sociology, anthropology, linguistics, among others. Its connotation is open to multiple interpretations and can be understood as broadly as the process of a social event (as indicated from its Latin origin *discursus*), and as narrowly as the connections among and across sentences in a piece of writing (as defined in some linguistic studies, cf. Gee 2014). Within language studies, the term "discourse" is further used in different ways. Some linguists (e.g. Halliday & Hassan, 1976) considered "discourse" similar to "text", referring both to written and

spoken language; some (e.g. Widdowson, 2007) preferred “discourse” for written language; and still some (e.g. Coulthard, 2014) preferred “discourse” for spoken language and “text” for written language. Due to different interpretations of “discourse”, multivariate terms cluster around upon touch of discourse features: cohesion, coherence, structure, organization, rhetorical patterns, style, register, genre and the like.

In assessing discourse features in writing, similar to the theoretical conceptualization, there is lack of consensus on features of similar discourse assessment constructs. Take for instance the first best-known rubrics in second language writing by Jacobs et al (1981). They assessed writing from five aspects: content, organization, vocabulary, language use, and mechanics. Constructs of content and organization are related to the key components of discourse features as conceived in assessment theory (Bachman, 1990). Although the same term is used, their interpretation of the assessment scope has both hierarchical overlaps and mismatches. According to Jacobs et al, organization involves fluent expression, clearly stated/supported ideas, logical sequencing and cohesion; content means relevant, knowledgeable, substantive and thorough development of thesis. Different from Jacobs et al (1981), Brown and Bailey (1984) designed an analytical scoring rubric for evaluation of classroom writing, in which organization constitutes introduction, body, and conclusion; and logical development of ideas embraces content. Here, Brown and Bailey’s scope of “organization” differs from that of Jacobs et al (1981); but together with other discourse elements within each set of rubric, these two rubrics share the same core discourse features of cohesion, organization and semantic relations as conceived in assessment theory.

Examination of second language assessment rubrics show that there are both overlaps and mismatches concerning understanding of discourse features. For a systemic and precise understanding of discourse level features, we further traced “discourse” from current linguistic theories. Among various schools of linguistic studies, systemic functional linguistics (SFL) provides a special advantage for teaching and learning guidance. SFL conceives language use into five levels of sub-systems (phonology, lexico-grammar, discourse semantics, register, and genre). These systems have three types of meta-functions (ideational, interpersonal, and textual). The systematic and functional perspective provides us a comprehensive framework for analyzing students’ discourse features in writing. Clarification on notion and scope of discourse features provides a sound basis for the purpose of diagnosing specific discourse problems in writing. Guided by the operational understanding of discourse constructs, we conducted further literature review on existing empirical studies to identify typical discourse problems in Chinese tertiary students’ English writing.

2.2. Studies on L2 Learners’ Discourse Features in Writing

Problems in Chinese tertiary students’ English writing have been examined extensively from a range of perspectives, both at global (comprehensive view of discourse features) and local levels (such as particular focus on conjunction features). Given the present emphasis on diagnosis of students’ specific problems in writing, related findings were categorized according to the key components of discourse features as conceived above.

2.2.1. Topic Decision

An effective writing begins with a single focus that is closely related to the writing topic. A frequent observation is that students have poor skills in identifying a clear focus or their perception of the rhetorical problem may be partially off-topic, especially for low-proficiency writers (e.g., Guo & Wang, 2005; Li, 2015; Liang, 2006). Some students may drift away from the topic by addressing only part of the writing prompt.

2.2.2. Clarity of Thesis Statement

The thesis statement declares the overall intention or controlling idea of the entire essay. One typical situation is that students’ thesis statement is unclear/too-general or non-existent and this is frequently considered as related to Chinese students’ confusing concept between “I” and “we” (e.g., Lee & Deakin, 2016; Wang, 2010; Yu, 2012). For instance, Wang (2010) asked students to choose one topic “*why I like to learn English*” or “*why I hate to learn English*” and many students have changed the subject “I” into “we/a person/you” (e.g. “*I think learning English can make our life more wonderful*” or “*English is a tool for people to communicate with each other*”).

2.2.3. Organization

Organization involves paragraphing of the writing as well as logical organization (Knoch, 2009). A typical discourse structure usually comprises five main elements: introduction (introductory information for lead in), thesis statement (the overall intention of the writer), main points (for supporting the thesis statement), supporting details (for supporting each main point) and conclusion (Burstein, 2009). These five elements work together to help present an organized logical discussion. It is found that some students tend to have loosely related structures in their writing (Yu, 2012) or produce ineffective introduction and conclusion (Chong et al, 2014). For instance, Yu (2012) observed that Chinese students may discuss very casually at the beginning of writing, with no obvious focus or idea progression pattern. Chong et al (2014) found that the top two errors at the discourse level were poor conclusion and poor introduction.

2.2.4. Content Development

Content development mainly means reiteration of the central ideas and full and/or sophisticated development of ideas. Many studies identified in students' writing the coherence break or unrelated idea progression in the connection of ideas in the passages (Hong & Xu, 2016; Wang & Sui, 2006; Yu, 2012). For instance, Yu (2012) found that sometimes students wrote a topic sentence at the beginning but failed to develop it. They may shift the topic without any purpose, which will affect the flow of information and hinder the logical connection.

2.2.5. Appropriate Use of Cohesive Devices

Cohesion is considered as the glue that holds the text together. Halliday and Hasan (1976) identified five typical linguistic components of cohesion: reference, substitution, ellipsis, conjunction, and lexical cohesion. In academic writing, conjunction, reference and lexical cohesion are the three most frequently used devices. Of these three typical cohesive devices, one frequent observation is absence or limited use of cohesive devices or logical connections among sentences and chunks (Ong, 2011; Liu & Guo, 2013; Lee et al, 2015). A second problem is misuse or inappropriate use of transition markers, which may lead to mismatch between the transition markers and the message communicated. The third typical problem is related to misuse of reference. For instance, Chan (2010) identified ambiguous use of "it" for reference was the top one discourse errors.

Synthesis of existing findings provides a systematic understanding about Chinese learners' specific discourse problems in English writing. However, the findings are based on study of different learner groups, some focusing on English majors and some on non-English majors, some on undergraduates studying in mainland China, and some on Chinese undergraduates studying overseas. To provide targeted instructions in the writing course, the present study conducted a systematic diagnostic assessment of first year undergraduate students' writing strengths and weaknesses at the beginning of their university study.

Our overarching research question is what problems these Year One undergraduate students have with their English writing. Three sub-questions are designed to guide the study: (1) What are students' average writing proficiency at the threshold of their university English writing? (2) What are their specific discourse problems in English writing? (3) What are the similarities and/or differences between high-graded and low-graded essays, between students from different majors, and between female and male students?

3. Materials and Methods

3.1. The Writing Context and Focal Participants

The present writing course is situated in a China-foreign

joint international educational program situated in a comprehensive university in mainland China. The annual enrollment of students in this program consist of about 600 students in 20 groups majoring in business, accounting, finance, aviation management and computer engineering studies.

The focal participants are Year One students from the finance and aviation management groups. First, they are chosen in that compared with their computer engineering and accounting peers, the finance and aviation management students are expected to have higher proficiency on communication skills and thus have higher motivation to improve their language and communication skills for their disciplinary study and future professional development. Second, Year One students are chosen in the expectation of gaining a comprehensive insight into students' writing ability at the beginning of their learning in university.

3.2. Data Collection and Data Analysis Methods

Altogether 66 pieces of writing on the same topic by two groups of students (33 from aviation management group and 33 from finance group) are gathered. These texts are the first relatively complete writing by these undergraduate students after their entry into university. These texts are graded along a set of 5-point scales in 16 categories in five main constructs: task fulfillment, logical progression of ideas, vocabulary, grammar, and cohesion. Specifically, the five constructs are developed for three aspects of consideration: typical constructs as used for classroom writing needs, theoretical conceptualization of discourse features, and common assessment requirements in academic writing. One main goal of the classroom learning is preparing students for IELTS writing. Thus, the present assessment criteria have its prototype framework with reference to IELTS writing requirement. Meanwhile, from daily teaching practice, we noticed that the coherence and cohesion part pose a great difficulty for students' writing. Hence, coherence and cohesion are further separated into two individual constructs. Besides, while assessing students' essays, we found some typical spoken features in students' writing. This did not fit the formal writing register. It is necessary to waken students' awareness on register. But given the target L2 learners' practical learning environment, the register features is not significant enough to be an independent construct itself. As a result, we added register to be a sub-skill of task achievement.

In total, we have 16 subskills under five main categories. The first category is task achievement, which includes [a] Genre appropriateness, [b] Register awareness, [c] Relevance and focus, and [d] Clarity of topic sentence or thesis statement. The second is Content and organization, which includes [e] Effective organization, [f] All main points directly supporting the topic sentence, [g] All supporting details directly support the main points, and [h]

Full or sophisticated development of ideas. The third is Lexical resources, which covers [i] Use of fresh words and expressions and [j] Collocation of words and expressions. The fourth is Grammar and mechanics, including [k] Variety of sentence patterns or structures, [l] Accuracy of sentence patterns or structures, [m] Grammatical accuracy, and [n] Spelling, capitalization, punctuations and other mechanics. The last category is Cohesion, covering [o] Appropriate use of conjunctive signals and [p] Appropriate use of other cohesive devices. These 16 features can be understood independent of the main constructs and in direct relation to overall writing quality. The final grade for each writing was achieved by computing the score of all the scales while incorporating the weighting convention as adopted in many other writing practices.

Two researchers graded nine trial essays first. Statistical results showed that the inter-rater reliability has a Spearman correlation coefficient of .857. We discussed the typical differences (features that have more than one scale's difference at each sub-construct) and agreed on a more consistent criteria. Two researchers continued to grade all the 66 pieces of writing. Marking of these 66 works by two raters resulted in a Spearman correlation coefficient of .839, which is acceptable in assessing overall writing quality according to Smagorinsky (2008). The final grade of each essay was two raters' average score. Non-parametric tests are performed to further analyze the significance of difference over the 16 sub-skills between different groups of learners. The following part reports some findings from the grading work.

4. Findings and Discussion

In this part, we first present students' overall writing ability. We continue to have closer examination of students' writing features by grouping them into high and low, within discipline, and between male and female students.

RQ: (1) *What are students' average writing proficiency at the threshold of their university English writing?*

In general, students' scores demonstrated negatively skewed distributions (mean = 84.93, standard deviation = 7.71, skewness = -1.044), as shown below in Table 1. Specifically, 32% students got scores 90 or above, 47% students got scores between 80~89, and 21% students got scores below 80.

Table 1. Description of student' overall writing proficiency

	Number	Mean	Standard deviation	Maximum	Minimum
Overall	66	84.932	7.7101	97.00	55.50
High-graded students	20	93.050	2.4165	97.00	90.00
Low-graded students	20	75.80	5.8544	83	56
Finance students	33	87.8182	6.1920	97	56
Aviation students	33	82.5455	8.2844	97	72
Female students	44	86.2045	8.1628	97	56
Male students	22	83.1364	6.4682	93	72

The high average score and the lower distribution toward the right-side tail are assumed reasonable in classroom context. First, classroom assessment is of low-stakes, intending to engage learners effectively in frequent practice. It has low stakes of creating significantly negative impact on their final grade. Second, classroom assessment is learning-oriented, giving priority to the process of learning. In the target context, before submitting their work for final grading, students had made rounds of revision based on teacher and peer feedback and other learning resources both in class and outside class. The ideal expectation was that by the final submission of this writing task, students would have removed as many grammar and vocabulary mistakes as detectable.

RQ2: (2) *What are the specific features in relation to discourse competence in students' English writing?*

Students' overall writing quality is assessed on 16 subcategories along a five-point scale. Their detailed performance on the 16 categories are shown in Table 2 below. Sorting these 16 categories according to their mean value, we got the top five categories *c* (4.735), *e* (4.625), *f* (4.561), *b* (4.561), and *n* (4.553), the middle six categories *d* (4.394), *g* (4.438), *o* (4.318), *p* (4.265), *h* (4.235), and *l* (4.212), and the bottom five categories *i* (4.114), *m* (4.083), *k* (4.083), *a* (4.038), and *j* (3.553). That is, students are good at presenting a relevant and specific focus (*c*) and a logical structure (*e*); most of their main points can directly support their topical statement (*f*); the language used are mostly in formal writing style (*b*); and most of them are aware of spelling, capitalization, punctuations and other mechanics features (*n*). In contrast, their relative weaknesses are collocation of words and expressions (*j*), basic awareness of discourse features (*a*), sentence variety (*k*), grammatical accuracy (*m*), and lexical range (*i*).

Table 2. Description of 66 student' writing festures

	N	Minimum	Maximum	Mean	Std. Deviation
a	66	2.0	5.0	4.038	1.4041
b	66	2.5	5.0	4.561	.4430
c	66	1.5	5.0	4.735	.6458
d	66	2.0	5.0	4.394	.7622
e	66	1.5	5.0	4.629	.7406
f	66	2.0	5.0	4.561	.7622
g	66	2.0	5.0	4.348	.6960
h	66	2.0	5.0	4.235	.7806
o	66	2.5	5.0	4.318	.6302
p	66	2.5	5.0	4.265	.7350
i	66	3.0	5.0	4.114	.5099
j	66	1.0	5.0	3.553	.8234
k	66	2.5	5.0	4.083	.5518
l	66	2.5	5.0	4.212	.5481
m	66	1.0	5.0	4.083	.8208
n	66	2.5	5.0	4.553	.5625

Specifically, among all the 16 categories, *a*, *b*, *c*, and *d* are related to task achievement; *e*, *f*, *g*, and *h* are related to content and organization; *o* and *p* cohesion; *i* and *j* to lexical resources; *k*, *l*, *m*, and *n* to grammar and mechanics. For task achievement, students are relatively good at *b* and *c*, weak at *a*, and medium at *d*. That is, students can present a relevant and specific focus, and follow the usual formal written convention; they are relatively weak at distinguishing particular features between paragraph and essay writing; and they can better improving their writing by clarifying their topical sentences with appropriate guiding information.

Regarding content and organization, students are relatively stronger at *e* and *f*, and medium at *g* and *h*. That is, they are good at organizing their ideas by following typical conventions and patterns of writing; and their main points can mostly support the topic or thesis statement. On the other hand, their supporting details may not be closely related to the main point and there may be need of more details to fully support the main point. What's more, students can better improve their writing by having sophisticated thinking on idea development, such as by making causal explanations or making comparisons if possible. On the linguistic side, in relation to cohesion, students in general remain medium. That is, students can use conjunction words and other cohesive devices with some fluency and accuracy, but they do not show much strength in employing various kinds of cohesive ties.

Concerning lexical resources, students show weaknesses

both on lexical range and appropriate collocation of words and expressions. Collocation in our study mainly refers to semantic collocation/correlation of words and expressions. It seems that students' lexical proficiency have a great impact on students' overall writing quality and insufficient vocabulary knowledge and skills may impede students' further improvement in writing.

For grammar and mechanics, students are good at handling mechanics features, remain medium at sentence accuracy, but are weak at sentence variety and grammatical accuracy. In other words, their sentence skills need much improvement and they also need to pay more attention to more detailed and specific grammatical features.

RQ: (3) *What are the similarities and/or differences between high-grade and low-grade essays, between finance and aviation management majors, and between female and male students?*

In order to gain deeper understanding about students' writing features and perceive possible influencing factors, we further grouped the 66 students respectively according to their overall writing proficiency (high and low), their disciplinary backgrounds and gender. The highest 20 writing and the lowest 20 ones were further selected for comparative analysis; there are the same number of students (33 for each) for both finance and aviation management program; the proportions between males and females are 22:44. Non-parametric tests are performed for each pair of group to perceive possible correlation. The comparative results are summarized in Table 3.

Table 3. Correlation among the 16 categories

Category	a	c	d	e	f	g	h	o	p	i	j	k	l	m
High-Low	√	√	√	√	√	√	√	√	√	√	√	√	√	√
Finance- Aviation			√	√	√			√	√		√			√
Male-Female							√			√				

Specifically, regarding high- and low-proficiency groups, the non-parametric tests on two independent samples show that there are significant differences on the 14 subcategories of writing. That is, the high-graded writing on average perform better than low-graded writing on all the 14 subcategories of writing. with only two exceptions: *b* (register) and *n* (mechanics). In particular, the top six strengths for high-graded writing are *a*, *e*, *c*, *f*, *d* and *p*, starting from the strongest; and their bottom three weaknesses are *j*, *i*, and *k*, starting from the weakest. Comparatively, the low-graded students have their top five strengths at *c*, *e*, *f*, *i* and *o*; and their bottom three weaknesses are at *j*, *a*, and *m*. Overall, high-graded students on average perform significantly better than low-graded students at 14 categories. Meanwhile, these two groups of learners have similar strengths at *c* (relevance and focus), *e* (effective organization) and *f* (All main points directly supporting the topic sentence); and share similar weaknesses at *j* (Collocation of words and expressions). Besides, high-graded students also have particular

problems with *i* (Use of fresh words and expressions) and *k* (Variety of sentence pattern); while low-graded students have particular problems with *a* (Genre appropriacy), *d* (Clarity of topic sentence), and *m* (Grammatical accuracy).

Concerning possible distinctive features between different disciplines, the non-parametric tests on two independent samples show that there are significant differences on seven subcategories of writing: *d*, *e*, *f*, *o*, *p*, *j*, and *m*. Finance students are on average perform better than aviation students in these seven categories. This result partly corresponds to the high- and low-graded results as discussed in previous part in that the finance students were recruited for entry into the program with a much fiercer competition. Consequently, students in the finance program were in general assumed to have higher English proficiency. Meanwhile, despite the overall better performance, finance students have similar writing patterns with aviation students in that the two groups' mean value follow the same order when sorted from the largest to the smallest. Both the mean value and the standard deviation shares similar pattern with the comparisons between high-graded and low-graded writing. Hence, it is likely that the significant differences between finance and aviation students on these seven subcategories are mainly related to these students' overall writing proficiency.

In respects of writing features by different genders, the non-parametric tests on two independent samples show that there are significant differences on two categories *h* and *i*. That is, female students on average perform better than male students on full or sophisticated development of ideas (*h*) and have wider lexical range (*i*). This is consistent with our daily teaching observation. In daily language teaching and learning practice, we observed that male students and female students have some different writing features; and in most situations, girls are likely to perform better in language learning than boys.

The overall analysis show that the most prevalent weakness in students' writing is lack of appropriate genre awareness (*a*). Students may be unaware of the difference between paragraph and essay writing, and furthermore, between personal writing and academic writing. This is understandable for Year One students at their beginning of university writing. As shown in China's Standard of English Ability (MOE, 2018), in the secondary school education, English learning has more emphasis on vocabulary accuracy and sentence fluency. While upon entry into tertiary level learning, there are more requirements on discourse features.

The next prevalent weakness is related to lexical usage (*j*) and syntactic skills (*m*). It is likely that L2 language proficiency may impede students from further communicating their genuine argument or reasoning in writing. An alternative explanation is that compared to subject teachers, language teachers may have more strict criteria in assessing vocabulary and grammar, as discussed in some studies (e.g. Greasley & Cassidy, 2010). Hence, it

is necessary to incorporate subject teachers' views when developing particular writing assessment criteria.

The third prevalent weakness is insufficient development of ideas. As demonstrated in categories *h* and *o*, many students only list examples to support their main points, with little intention for causal or comparison argument. Similar to Deng's (2006) observation, students prefer to highlight the quantity of their arguments to support their standpoints. The prevalence of general discussion or insufficient development of idea may also be related to the length requirement on students' writing (≥ 100 words). Hence, it is critical to accurately identify the underlying cause or contributing factors for each feature and devise meaningful development strategies accordingly.

Furthermore, students can better improve their writing quality by clarifying their topical or thesis statement (*d*). Many students wrote their thesis in very general sense, showing little personal commitment and/or specific guide for following discussion. One frequently suggested reason is cultural influence, which is further related to indirect thinking and collective thinking. Chinese students are found unable to distinguish individual voice from collective voice, which may create confusion by considering collective voice as similar to individual stand. Holding such a writing habit, students tend to delay or somewhat neglect the significance of thesis statement. The weakness in thesis statement may also indicate that students have not spent enough time planning their writing. They used to begin their writing upon seeing the topic on hobby and take that it will be enough as long as they develop their writing on something related to hobby. A further reason may be related to students' motivation in writing. They may write just for fulfilling the writing task or meeting the needs of a test and do not have a particular purpose for communicating their genuine ideas. However, in the academic context, to produce an effective writing, it is more important to keep the audience in mind and conveys something worthwhile and interesting to the audience. As is observed in some studies (e.g. Chen 2016), it is widely acknowledged among students that the only purpose for English writing is for test preparation. To address this situation, it is critical to awaken students' awareness of other key purposes of writing and arouse their internal motivation in English writing.

5. Conclusions

With the guidance of core diagnostic assessment principles, this study develops a set of fine-grained writing assessment instruments and provides a coherent picture of syntactic, lexical, rhetorical and other discourse features in Chinese first year students' English writing. Overall, the target students demonstrate a relatively good mastery of overall writing ability, performing *good* in most of the targeted writing constructs. Meanwhile, the most prevalent

weakness is related to overall discourse awareness, vocabulary and sentence proficiency, and full and/or sophisticated idea development. Based on knowledge of students' specific strengths and weaknesses, more targeted remediation prescription can be designed and delivered for maximal support in facilitating Chinese undergraduate students' development of English writing in classroom setting.

Nevertheless, the present study only assesses students' writing performance in addressing one writing prompt. It is worthwhile exploring whether there would be similar strengths and weaknesses patterns using a different writing prompt. Besides, the present analysis of students' writing is based on grading of writing with a tentative assessment rubric. While this may be informative in conveying an overall picture about students' writing ability, more work is needed to elicit detailed evidence and further explore potential influential factors.

Acknowledgements

The authors would like to thank International Education College at Henan University and Department of Linguistics and Modern Languages at the Education University of Hong for providing valuable researching sources and support. They are also grateful to the anonymous reviewers and the editors for their critical and constructive suggestions to improve this article.

Funding

The study described in this article was partially supported by a research project [grant number # 2019CXD007] from Henan University. It would also like to acknowledge the support of two research projects (SKL-2015-1349 and SKL-2019-1560) from Social Science Association of Henan Province and the support (2015JKGHYB0058) from the Education Department of Henan Province.

REFERENCES

- [1] Alderson, J. (2005). Diagnostic foreign language proficiency: The interface between learning and assessment. New York: Continuum.
- [2] Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- [3] Brown, J., & Bailey, K. (1984). A categorical instrument for scoring second language. *Language Learning*, 34(4), 21-38.
- [4] Burstein, J. (2009). Opportunities for natural language processing research in education. *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 6-27). Heidelberg: Springer.
- [5] Chan, A. (2010). Toward a taxonomy of written errors: Investigation into the written errors of Hong Kong Cantonese ESL learners. *Tesol Quarterly*, 44(2), 295-319.
- [6] Clarke, B. (2017). Cohesion in systemic functional linguistics: A theoretical reflection. In T. Bartlett, & G. O'Grady, *The Routledge Handbook of Systemic Functional Linguistics* (pp. 404-417). Oxon: Routledge.
- [7] Chen, Q. (2016). Reconstruction of the teaching model of college English writing in the Big Data era. *Foreign Language Research*, 190(3): 129-132.
- [8] Chong, K. (2014). Automated prediction of the written errors of tertiary level ESL and EFL learners. *Arizona Working Papers in SLA & Teaching*, 84-109.
- [9] Coulthard, M. (2014). *An Introduction to Discourse Analysis*. New York: Routledge.
- [10] Deng, Y. (2006). Use of Adverbial Connectors in Chinese EFL Learners' Argumentative Writing. *CELEA Journal*, 29(6): 32-36.
- [11] Gee, J. (2014). *An Introduction to Discourse Analysis: Theory and Method*. Fourth ed. London: Routledge.
- [12] Pete Greasley & Andrea Cassidy (2010) When it comes round to marking assignments: how to impress and how to 'distress' lecturers *Assessment & Evaluation in Higher Education*, 35(2), 173-189,
- [13] Guo, L., & Wang, H. (2005). Analysis of error types in Chinese English learners' writing. *Sina-US English Teaching*, 9-13.
- [14] Halliday, M., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- [15] Hong, M., & Xu, Y. (2016). A Study on Local Coherence Characteristics of Non-English Majors' English Essays. *TEFLE*, 59-64.
- [16] IELTS. (2018, July 26). Retrieved from <https://www.ielts.org/teaching-and-research/test-statistics>
- [17] Jacobs, H., Zinkgraf, S., Wormuth, D., Hartfiel, V., & Hughey, J. (1981). *Testing ESL composition: A practical approach*. Rowley: Newbury House.
- [18] Knoch, U. (2009). Diagnostic writing assessment: The development and validation of a rating scale. Frankfurt: Peter Lang.
- [19] Lee, J., & Deakin, L. (2016). Interactions in L1 and L2 undergraduate student writing: Interactional metadiscourse in successful and less-successful argumentative essays. *Journal of Second Language Writing*, 21-34.
- [20] Lee, J., Yeung, C., Zeldes, A., Rezicek, M., Ludeling, A., & Webster, J. (2015). CityU corpus of essay drafts of English language learners: a corpus of textual revision in second language writing. *Language Resources and Evaluation*, 659-683.
- [21] Li, H. (2015). A Rhetorical Structure Theory Based Approach to the Erroneous Rhetorical Relations in Chinese EFL Learners' Argumentative Writing. *Science & Technology Vision*, 52-53.
- [22] Liang. (2006). A Study of Coherence in EFL learners'

written production. *Modern Foreign languages*, 284-292.

- [23] Liu, J., & Guo, P. (2013). An analysis of the incoherence and in cohesion in Undergraduate English writing. *Journal of Shaanxi Normal University*, 153-156.
- [24] McMillan, J. (2014). *Classroom assessment: Principles and practice for effective standards-based instruction* (Sixth ed.). Boston: Pearson.
- [25] Ong, J. (2011). Investigating the Use of Cohesive Devices by Chinese EFL Learners. *Asian EFL Journal*, 42-65.
- [26] Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic measurement: theory, methods, and applications*. New York: The Guilford Press.
- [27] Smagorinsky, P. (2008). The method section as conceptual epicenter in constructing social science research reports. *Written Communication*, 389-411.
- [28] Wang, X.W. (2010). TP Pattern and Coherence in English Writing: Analysis of TEM-4 Writing papers. *Foreign Language Research*. 153(2): 103-106.
- [29] Wang, H., & Sui, D. (2006). Measuring Coherence in Chinese EFL Majors' Writing through LSA (Latent Semantic Analysis). *Asian EFL Journal*, 1-24.
- [30] Widdowson, H. (2007). *Discourse Analysis*. Oxford: OUP.
- [31] Yu, A. (2012). Analysis of the Problems of the Chinese College Students' EFL. *International Education Studies*, 199-203.