English Language Teaching; Vol. 17, No. 2; 2024
ISSN 1916-4742 E-ISSN 1916-4750
Published by Canadian Center of Science and Education

Investigating the Effects of Dynamic Assessment on Chinese Undergraduates' English Writing Performance in the Blended Learning Context

Yawei Xiao[1]

[1] Foreign Language Colleges, Jiangxi Normal University, Jiangxi, China

Correspondence: Yawei Xiao, Jiangxi Normal University, Jiangxi, China. E-mail: 785391970@qq.com

Received: January 9, 2024 Accepted: January 26, 2024 Online Published: January 27, 2024

doi: 10.5539/elt.v17n2p37 URL: https://doi.org/10.5539/elt.v17n2p37

Abstract

Whereas the effectiveness of dynamic assessment has been investigated in multiple contexts, it has been under-investigated in the context of blended learning mode for writing performance. To address this gap, this study aims to explore the effects of dynamic assessment on Chinese undergraduates' writing performance as overall writing performance and writing complexity, accuracy and fluency in the blended learning context. To this end, a quasi-experiment was carried out with two intact classes from a Chinese university, one being the control group (n=34) and the other experimental group (n=36). A 12-week intervention was conducted in English writing classes under the blended learning mode, with the experimental group receiving dynamic assessment while the control group having traditional static assessment. At the end of the experiment, six students attended a semi-structured interview. The findings revealed that the experimental group improved significantly in writing performance in terms of overall scores, lexical density, lexical sophistication, and accuracy. However, dynamic assessment had no significant effect on lexical diversity, syntactic complexity and fluency. Besides, the interview findings evidenced that the students held positive attitudes toward the use of dynamic assessment in English writing classes in the blending learning context. Implications for writing instruction and future research are discussed.

Keywords: dynamic assessment, blended learning, English writing performance, complexity, accuracy, fluency

1. Introduction

Shortly after the Internet became available to a wide variety of nations, traditional teaching methods underwent a shift into modern, online, internet-mediated approaches to education (Bernard et al., 2009). Taking advantage of both offline and online learning to increase teaching efficiency and improve learning outcomes, blended learning has emerged as a significant educational trend. Previous studies on blended learning offer valuable insights into the merits of this approach to learning. These include enhancements in student outcomes, increased flexibility for educators and learners, heightened accessibility to education, and the accommodation of individualized student needs (Dakduk et al., 2018; Yusoff et al., 2017).

Dynamic assessment (DA) has sharply drawn the attention of second language (L2) scholars and teachers in recent decades since it is essential for the development of L2 learners' skills. Poehner (2008) notes that it is an alternative assessment method to the static methods, emphasizing that assessment and teaching are fully integrated. DA originates from Vygotsky's (1978) sociocultural theory (SCT), with its core of the Zone of Proximal Development (ZDP), which is conceptualized as the distance between what an individual can do independently and what he can complete under the assistance of adults or more capable peers (Vygotsky, 1978). DA integrates assessment and instruction in an interactive and supportive context to move toward an emergent future. Both highly experienced experts and peers could work together to construct a scaffolding ZDP for learners through socially collaborative activities, referred to as mediation (Lantolf & Poehner, 2011). Based on the principles of DA, it sets out to explore the learners' level of independent mental development and their potential levels of development.

Serving as an alternative to standardized testing, DA has been applied in many classroom teaching settings and its effectiveness on learners' language development has been substantiated by researchers (Antón, 2009; Mauludin, 2018; Rashidi & Bahadori Nejad, 2018). Blended learning combines face-to-face instruction with

online learning, which implements targeted interventions in the DA (Qian et al., 2023). There seems to be scant attention to the dynamic assessment embedded in the blended learning context. To address this gap, this study aimed to explore the effects of DA on Chinese undergraduates' English writing performance in terms of overall scores and writing complexity, accuracy and fluency in the blended learning context.

2. Literature Review

The efficiency of DA in developing learners' L2 writing abilities has been highlighted by a plethora of empirical research (Antón, 2009; Rahimi et al., 2015). An analysis of the literature highlighting its contributions and the gaps that need to be filled in further research is provided below.

2.1 Studies of Effects of DA on Learners' Writing Performance

A mass of empirical studies has been carried out to explore the impact of DA on L2 writing (Antón, 2009; Kushki et al., 2022; Rahimi et al., 2015). One of the earliest attempts was conducted by Antón (2009), who ran a five-part diagnostic test in a university language program. Results revealed that DA enabled mediators to understand learners' actual and emergent abilities and programs to devise individual instructional plans based on learners' needs. Even though the DA writing procedures were carried out only once and did not include any specifics, this study provided illumination and inspiration for later studies.

Other studies have compared the effects of dynamic assessment and static assessment to testify to its effectiveness, most of which have reached a consensus that dynamic assessment is more efficient. For instance, following the dynamic assessment, Shrestha and Coffin (2012) investigated the efficacy of tutor mediation in academic writing within the context of undergraduate business students engaged in open and distance learning. Their findings indicated that DA has the potential to enhance the academic writing development of undergraduates by addressing their specific needs. Students enjoyed DA procedures and perceived them to be supportive of their writing performance. In the same vein, the study conducted by Hadidi (2023) also provided evidence for more significant advantages of DA than that of static assessment on learners' ability to generate written argumentative discourse.

As computer-based language assessments have proliferated, scholars have inclined their focus on applying DA with technology-aided tools. For instance, Tzuriel and Shamir (2002) conducted a comparative analysis of the impacts of DA in two distinct treatment groups of young children. One group received computer-assisted mediation, while the other received examiner-only mediation. Findings showed that computer-assisted mediation achieved considerable cognitive changes compared to face-to-face mediation. However, Ebadi and Rahimi (2019) investigated the impact of online dynamic assessment on EFL learners through one-on-one individual and online synchronous DA sessions in Google Docs. It explored the short and delayed effects of DA on writing performance by conducting near and delayed transfer tasks. The results of the microgenetic development of learners indicated the learners' regressed academic writing performance in the far transfer task. Interview analysis highlighted learners' positive perceptions of the effectiveness of DA.

Although a few studies have documented that DA positively impacts students' writing performance, scant attention is given to how it may affect the blended learning context. Through an empirical study, Qian et al. (2023) explored the effectiveness of DA-embedded SPOC blended teaching in promoting students' development of paraphrasing concepts and skills. The model helps students correct misconceptions and promote their paraphrasing skills, but there is an imbalance in the development of paraphrasing skills.

In summary, despite the growing quantity of DA research in writing instruction, there are still some lacunae in this field. Firstly, nearly all the studies were conducted in traditional or online contexts, and only a few scholars have investigated DA in the blended learning context. In addition, other investigations of the effectiveness of DA assisted by technological tools have obtained mixed results.

Motivated by these gaps, this study aims to examine the effect of dynamic assessment on Chinese undergraduates' writing performance in the blended learning context. These research questions were put forward:

(1) What are the effects of dynamic assessment on Chinese undergraduates' English writing performance in terms of overall scores in the blended learning context?

(2) What are the effects of dynamic assessment on Chinese undergraduates' English writing performance in terms of writing CAF in the blended learning context?

(3) What are the students' perceptions of dynamic assessment in the blended learning context?

3. Methodology

3.1 Participants and Setting

A group of 70 second-year English major undergraduates from a Chinese university were selected as participants by running a convenience sampling method, which involves choosing participants who are convenient to researchers (Riazi, 2016). They were from two intact classes attending an English writing course taught by the same teacher. These classes were designated the EG (n=36) and the CG (n=34). In both groups, blended learning was applied in the English writing courses, which consisted of a cohesively offline learning session in the traditional classroom and online learning sessions based on the iWrite 3.0 system. The EG underwent a dynamic assessment, whereas the CG received a traditional assessment. To check participants' homogeneity regarding prior written proficiency, an argumentative writing task serving as a pre-test, similar to TEM-4 (Test for English Majors-Band 4), was administered to both groups, and essays were rated by two experienced teachers from this university using the same rubric with 20 maximum total points (see Appendix). It was found that writing proficiency between the groups was not significantly different before embarking on the study ($p= 0.968 > .05$).

3.2 Instruments

3.2.1 Writing Tasks

Five writing tasks were involved in this research. Two served as the pre-test and the post-test; the others were employed for writing training during the intervention. In line with the requirements of the writing section in TEM-4, the participants were asked to complete the texts in 35 minutes, with at least 200 words required. Two experienced teachers in this university rated both tests with the same holistic rubric. Besides, inter-rater reliability in the test scores was calculated using Pearson correlation analysis. The inter-rater reliability coefficients were 0.777 ($p=0.000 < .05$) for the pre-test and 0.797 ($p=0.000 < .05$) for the post-test, indicating the relatively high consistency between raters.

3.2.2 Interview

A semi-structured interview was another instrument employed in the study, mainly modified from Zhang's (2012) study. The interview was run in their mother tongue, Chinese. Each interview lasted 20 to 30 minutes, depending on how long each interviewee needed to share their perspectives. All the interview data were audio-recorded under the interviewees' consent, transcribed verbatim, and translated into English.

3.2.3 IWrite

iWrite 3.0 system, developed by Foreign Language Teaching and Research Press, is an online platform with writing instruction and an automatic writing evaluation system. It not only provides holistic scores but also reports on linguistic features. Students can examine the evaluation of the relevance and coherence of the content as well as alternative expressions offered by the platform when they check their essays. In addition, it also offers the function of organizing instruction and learning and the function of peer assessment at any time. The teacher can assess each student's essay in the database, after which they can deliver timely, individualized training and learning performance analysis. Based on the iWrite 3.0 system, a blended learning environment is developed, integrating offline and online writing sessions.

3.3 Procedures

The research lasted for 14 weeks, from March to June 2023. Initially, a pre-test was administered in two groups to examine their English writing proficiency, while the post-test was conducted after the 12-week intervention.

During the intervention, the EG and CG received different assessment models. DA was applied in the EG, while the CG adopted the traditional assessment. The DA model in the blended learning context is illustrated in Figure 1. It includes three sub-stages: pre-writing, while-writing and post-writing. In the pre-writing stage, the teacher assigned tasks and requirements for the writing topic before class. The students then conducted online learning activities in the iWrite 3.0 system. In the class, the teacher assigned writing topics, organized brainstorming activities, and delivered writing strategies. Next, students were required to complete their first draft independently and submit it to the iWrite 3.0 system.

Then, the teacher modeled how to give peer assessment. Being illustrated the peer feedback guidance sheet adapted by Min (2005) and Yang et al. (2006) in detail, students became familiar with the criteria for feedback, which mainly focus on content, organization, vocabulary, grammar and mechanics. After class, students anonymously provided feedback on each other's writings through the iWrite 3.0 system. After the peer assessment, the teacher conducted modified strategies training and provided writing revision strategies for students in the class. After revising their first drafts based on peer assessment, students submitted their second

drafts to the iWrite 3.0 system for teacher's assessment. Finally, according to the teacher's feedback, students modified their final drafts to complete them. At the last stage, the teacher established a sample of excellent compositions on the online platform for students to appreciate, and he also elaborated on the model essays in the class.

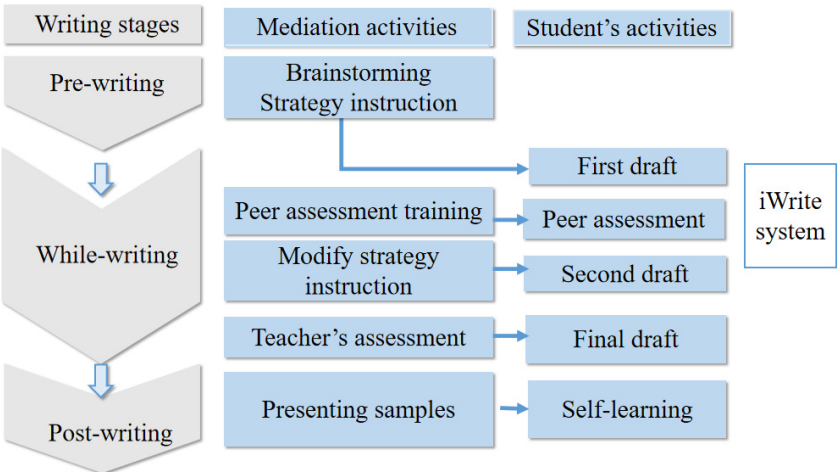

Figure 1. Model of Dynamic Assessment Teaching Procedures

The detailed writing teaching procedures in the CG are as follows (see Figure 2). Like the EG, students in the CG received writing prompts and necessary guidance on brainstorming and writing strategies before writing. They were required to write independently and submit their compositions to the iWrite 3.0 system. After writing, the teacher scored their first drafts with brief comments on the online platform. Model essays were also selected and shared for students' reference.

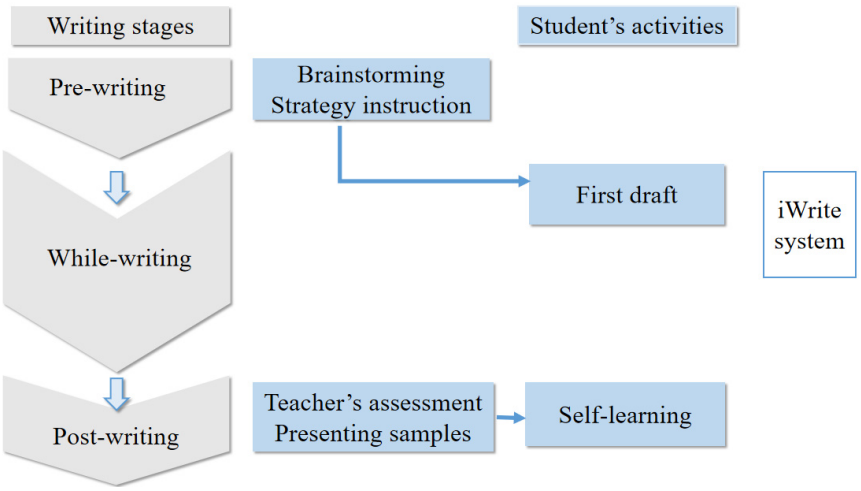

Figure 2. Model of Traditional Assessment Teaching Procedures

After implementing the post-test, the semi-structured interviews were conducted with six students invited.

3.4 Data Collection and Analysis

To address the three research questions, a mixed method was utilized through triangulation of the data, which allowed researchers to get a fuller understanding of the multifaceted nature of a phenomenon (Mackey & Gass, 2016). The quantitative data included pre-and post-test scores and the measures of writing CAF of written texts. Two experienced teachers in this university were invited to mark all tests according to the same rating rubric (see Appendix). The ultimate score for these two tests was the average score provided by two raters.

In addition, texts were investigated in terms of the measures of CAF. Since past studies have confirmed the reliability of the automated tools for the coding of relevant measures (Polio & Yoon, 2018), all the essays were analyzed by utilizing computational tools, Lexical Complexity Analyzer (LCA) and L2 Syntactic Complexity Analyzer (L2SCA), developed by Lu (2010). Language complexity refers to lexical complexity and syntactic

complexity. Lexical complexity was further operationalized regarding three indexes: lexical density (LD), lexical sophistication (LS1) and lexical diversity (MSTTR) (Ellis & Yuan, 2004; Lu, 2010). The percentage of dependent clauses compared to the total number of clauses (DC/C) and the proportion of clauses to T-units(C/T) was obtained to gauge syntactic complexity (Wigglesworth & Storch, 2009). Accuracy was calculated by the number of error-free T-unit tokens divided by the total number of T-units (EFT/T) (Polio & Shea, 2014). Average word production words per minute (W/M) was used to assess fluency (Chenoweth & Hayes, 2001). Table 1 summarizes the measures of CAF in this study.

Independent sample *t*-tests and paired sample *t*-tests were performed utilizing SPSS 26.0 to assess the effects of DA on the learners' writing performance in the blended learning context.

Table 1. A summary of CAF measures

Measure types		Measures
Complexity	Lexical complexity	Number of words divided by the total number of words (LD)
		Types of lexical sophisticated words divided by the total amount of word types (LS1)
		Mean TTR of all 50-word segments (MSTTR)
	Syntactic complexity	Number of clauses divided by number of T-unit (C/T)
		Number of dependent clauses divided by number of clauses (DC/C)
Fluency		Words per minute (W/M)
Accuracy		Total number of error-free T-units divided by total number of T-units (EFT/T)

Concerning the qualitative part, a total of 6 students with 2 students from each of the three levels (high, medium and low levels) of writing proficiency according to the pre-test scores, were chosen from the EG to engage in the interview. The interviewees were chosen for the aim based on the criteria of purposive sampling (Denscombe, 2007). According to the pre-test scores, students in the EG were divided into three levels: high, medium and low. Those students whose scores ranked at the top 33.3 % of the whole class were distributed into the high-level group and those whose scores account for the medium 33.4% of the class were treated as members of the medium-level group. The rest of the students served as members of the low-level group. The written consent was provided for the interviews. Moreover, pseudonyms were employed to guarantee the privacy of respondents (i.e., S1, S2, … and so on). Thematic analysis, as outlined by Braun and Clarke (2006), was employed to analyze the interview data. The analytical process generally contains six steps: transcribing and familiarizing with the data, generating initial codes, generating themes, reviewing themes, naming and interpreting themes and producing reports. Accordingly, the author developed the original codes from the data with another postgraduate majoring in applied linguistics, and the consistency reached over 90%. Then, the major themes were formulated based on these codes and were grouped into main categories. At last, the individuals interviewed reviewed the excerpts and suggested revisions if there were discrepancies with their intended meanings.

4. Results

4.1 Results of Overall Scores

The first research question explored the effects of DA on students' writing performance, as indicated by overall scores. Table 2 and Table 3 display the results of the independent samples t-tests.

Table 2. Independent samples t-test for overall scores in the pre-test

EG (n=36)		CG (n=34)		t	df	p (2-tailed)
M	SD	M	SD			
13.806	1.0973	13.794	1.3034	0.040	68	0.968

Table 3. Independent samples t-test for overall scores in the post-test

EG (n=36)		CG (n=34)		t	df	p (2-tailed)
M	SD	M	SD			
15.125	1.1673	14.176	1.1735	3.389	68	0.001*

* Statistically significant difference: p < .05

According to Table 2, regarding the pre-test writing scores, the mean scores were 13.806 and 13.794, respectively, in the EG and CG, with the Sig (2-tailed) being 0.968 (p＞.05), which showed that the writing proficiency in the two groups was roughly the same at the outset of the study. As indicated in Table 3, the average scores of the EG (M=15.125, SD=1.1673) were higher than those of the CG (M=14.176, SD=1.1735). In addition, it demonstrated that a significant difference existed between the overall scores of post-tests of the two groups (p=0.001<.05). It could infer that the EG outperformed the control group regarding overall writing performance after applying DA.

Table 4 reports the results of the paired sample t-tests of the two groups' overall pre-and post-test scores. The descriptive statistics in Table 4 show that for the EG, the mean overall writing scores increased from 13.804 in the pre-test to 15.125 in the post-test. The mean scores of overall scores increased significantly (p=0.000＜.05). It revealed that a significant difference was observed between the two tests of the EG concerning overall scores. In other words, DA exerted a positive effect on improving students' overall writing performance. Concerning the CG, the mean scores of post-tests increased by 0.382 points with no significant difference between the two tests, as shown in Table 4, where Sig (2-tailed) equaled 0.074, more than 0.05, implying that the overall writing ability in the CG did not significantly improve.

Table 4. Paired samples t-tests of overall scores in the pre-and post-test for the EG and CG

Group	Pre-test		Post-test		t	df	p (2-tailed)
	M	SD	M	SD			
EG	13.806	1.0973	15.125	1.1673	-5.740	35	0.000*
CG	13.794	1.3034	14.176	1.1735	-1.848	33	0.074

* Statistically significant difference: p < .05

4.2 Results of CAF

The study further investigates students' writing performance from the perspective of writing CAF. Accordingly, follow-up analyses were performed by running independent sample t-tests and paired sample t-tests. Table 5 displays the results of three sub-components of the lexical complexity.

Table 5. Independent sample t-test for lexical complexity in the pre-test

Measure	EG (n=36)		CG (n=34)		t	df	p (2-tailed)
	M	SD	M	SD			
LD	0.4811	0.0324	0.4668	0.0291	1.945	68	0.056
LS1	0.2186	0.0452	0.2050	0.0575	1.104	68	0.273
MSTTR	0.7764	0.0367	0.7644	0.0372	1.357	68	0.179

Based on Table 5, the two groups had no significant differences for all measures of lexical complexity, LD (p=0.056 >.05), LS1 (p=0.273 >.05), MSTTR (p=0.179 >.05), which indicated that the two groups' lexical complexity was quite equivalent before the intervention.

Table 6. Independent sample t-test for lexical complexity in the post-test

Measure	EG (n=36)		CG (n=34)		t	df	p (2-tailed)
	M	SD	M	SD			
LD	0.5314	0.0349	0.5079	0.0389	2.657	68	0.01*
LS1	0.2722	0.0579	0.2438	0.0421	2.334	68	0.023*
MSTTR	0.7778	0.0354	0.7721	0.0342	0.687	68	0.494

* Statistically significant difference: p < .05

Table 6 shows that the EG had slightly higher mean scores in all indices of lexical complexity than the CG, among which a statistically significant difference was detected both in LD (p=0.01＜.05) and LS1 (p=0.023 ＜.05). According to the data of MSTTR, the difference between the groups is not statistically significant (p=0.494＞.05). The results suggested that the EG excelled the CG in the lexical complexity in terms of lexical density and lexical sophistication in the post-test.

Table 7. Paired samples t-test for lexical complexity in the pre-and post-test for the EG and the CG

Measure	Group	Pre-test		Post-test		t	df	p (2-tailed)
		M	SD	M	SD			
LD	EG	0.4811	0.0324	0.5314	0.0349	-6.082	35	0.000*
	CG	0.4668	0.0291	0.5079	0.0389	-5.178	33	0.000*
LS1	EG	0.2186	0.0452	0.2722	0.0579	-5.732	35	0.000*
	CG	0.205	0.0575	0.2438	0.0421	-3.183	33	0.003*
MSTTR	EG	0.7764	0.0367	0.7778	0.0354	-0.166	35	0.869
	CG	0.7644	0.0372	0.7721	0.0342	-0.865	33	0.393

* Statistically significant difference: p < .05

Table 7 reports the results of comparing the lexical complexity for pre-test and post-test in each group. The mean scores of LD, LS1 and MSTTR improved in the EG after the intervention. However, not all measures of lexical complexity were significantly increased, among which the average scores of LD and those of LS1 significantly differed in the two tests (p=0.000 < .05). No significant improvement was observed in the MSTTR as the Sig (2-tailed) was 0.869, higher than 0.05. The results indicated that the proportion of content word types and sophisticated word types in the EG was significantly higher after the treatment. The same is true for the mean scores of LD and LS1 for the CG since Sig (2-tailed) was 0.000 and 0.003, respectively, less than 0.05. However, the degrees of variation were not as great in the CG compared with the EG. In this sense, compared with the traditional assessment, DA was more effective in facilitating the lexical complexity, especially the lexical density and lexical sophistication. However, DA did not impact positively on students' lexical diversity.

To probe into the effects of DA on syntactic complexity, a comparison analysis between the two groups was conducted. Table 8 and Table 9 describe the results of the independent sample *t*-tests in the pre-test and post-test.

Table 8. Independent sample t-test for syntactic complexity in the pre-test

Measure	EG (n=36)		CG (n=34)		t	df	p (2-tailed)
	M	SD	M	SD			
C/T	1.5854	0.2459	1.5811	0.3452	0.06	68	0.952
DC/C	0.3672	0.0979	0.3469	0.1260	0.752	68	0.455

Table 9. Independent sample t-test for syntactic complexity in the post-test

Measure	EG (n=36)		CG (n=34)		t	df	p (2-tailed)
	M	SD	M	SD			
C/T	1.6061	0.2716	1.5949	0.2578	0.178	68	0.860
DC/C	0.3735	0.0994	0.3548	0.0839	0.849	68	0.399

As indicated in Table 8, the EG gained relatively similar mean scores of C/T and those of DC/C with the CG. Meanwhile, the results of independent sample t-tests did not demonstrate any significant differences in the syntactic complexity between the two groups before the intervention (p=0.952＞.05; p=0.455＞.05). They were equally similar in the writing performance in terms of syntactic complexity at the outset of the study.

Table 9 illustrates that the average scores for both the C/T (M=1.6061, SD=0.2716) and DC/C (M=0.3735, SD=0.0994) of the EG were slightly higher than those of the CG (M=1.5949, SD=0.2578; M=0.3548, SD=0.0839). However, the measures for syntactic complexity denoted that no significant differences for neither C/T (p=0.86＞.05) nor for DC/C (p=0.399＞.05) were observed. It was revealed that the application of DA had no significant impact on learners' syntactic complexity.

Using a paired samples t-test, Table 10 compares the pre-test and post-test in each group. The syntactic

complexity of the EG rose slightly in the CT, increasing from 1.5854 to 1.6061, and in the DC/C, rising from 0.3672 to 0.3735. Nonetheless, the results suggested that no significant differences between the pre-and post-test were detected in neither the C/T (p=0.686＞.05) nor in the DC/C (p=0.726＞.05). The same pattern was also found for syntactic complexity with no significant more means of C/T (p=0.836＞.05) and means of DC/C (p=0.711＞.05) in the CG. In this sense, DA failed to improve syntactic complexity.

Table 10. Paired samples t-test for syntactic complexity in the pre-and post-test for EC and CC

Measure	Group	Pre-test		Post-test		t	df	p (2-tailed)
		M	SD	M	SD			
C/T	EG	1.5854	0.2459	1.6061	0.2716	-0.407	35	0.686
	CG	1.5811	0.3452	1.5949	0.2578	-0.209	33	0.836
DC/C	EG	0.3672	0.0979	0.3735	0.0994	-0.354	35	0.726
	CG	0.3469	0.1260	0.3548	0.0839	-0.373	33	0.711

As for accuracy, the data from two tests were compared and analyzed, employing independent samples t-tests. Table 11 and Table 12 show the results of the accuracy measures in the pre-test and post-test.

Table 11. Independent sample t-test for accuracy in the pre-test

Measure	EG (n=36)		CG (n=34)		t	df	p (2-tailed)
	M	SD	M	SD			
EFT/T	0.5739	0.1689	0.5691	0.1425	0.128	68	0.898

Table 12. Independent sample t-test for accuracy in the post-test

Measure	EG (n=36)		CG (n=34)		t	df	p (2-tailed)
	M	SD	M	SD			
EFT/T	0.6916	0.1428	0.5941	0.1615	2.68	68	0.009*

* Statistically significant difference: $p < .05$

As depicted in Table 12, the average scores of EFT/T in the pre-test are almost identical for both groups. The Sig (2-tailed) was 0.898, more than 0.05, signifying no significant difference in participants' writing performance regarding accuracy. In other words, before the treatment, the EG and the CG performed similarly in accuracy. According to Table 12, the mean scores of EFT/T (M=0.6916, SD=0.1428) in the EG were higher than those of the CG (M=0.5941, SD=0.1615), implying a gap of 0.0975. Moreover, inspections of the two groups' means revealed that the average scores of EFT/T in the EG were significantly greater than in the CG (p=0.009＜.05), indicating that the EG produced fewer mistakes than the CG after the intervention.

Table 13. Paired samples t-test for accuracy in the pre-and post-test for EC and CC

Measure	Group	Pre-test		Post-test		t	df	p (2-tailed)
		M	SD	M	SD			
EFT/T	EG	0.5739	0.1689	0.6916	0.1428	-4.119	35	0.000*
	CG	0.5691	0.1425	0.5941	0.1615	-0.902	33	0.374

* Statistically significant difference: $p < .05$

Results of the paired samples t-tests are presented in Table 13, which displayed that the mean scores of EFT/T in the EG in the post-test (M=0.6916, SD=0.1428) were higher than those in the pre-test (M=0.5739, SD=0.1689). Moreover, the difference between the two tests was significant (p=0.000＜.05), indicating students' significant enhancement in their competence to produce more accurate writing. For the CG, the mean scores rose as well. Nonetheless, from Table 13, the Sig (2-tailed) turned out to be 0.374, greater than 0.05. No significant difference between the accuracy index for the CG in the two tests was detected. As a result, EG made greater gains in writing accuracy in a statistically significant way due to the implementation of DA.

Concerning fluency, the results of independent sample t-tests in the two tests are presented and analyzed as

follows.

Table 14. Independent sample t-test for fluency in the pre-test

Measure	EG (n=36)		CG (n=34)		t	df	p (2-tailed)
	M	SD	M	SD			
W/M	6.7937	1.2331	6.5319	1.1774	0.907	68	0.368

Table 15. Independent sample *t*-test for fluency in the post-test

Measure	EG (n=36)		CG (n=34)		t	df	p (2-tailed)
	M	SD	M	SD			
W/M	6.8008	0.7574	6.7849	1.0459	0.073	68	0.942

Table 14 shows that the mean scores of W/M achieved by the EG were relatively equal to those of the CG in the pre-test, which suggested that these groups did not differ significantly concerning the production of writing fluency of the pre-test ($p=0.368>.05$). In addition, Table 15 exhibits that no significant difference in the mean score of W/M between these two groups in the post-test was found ($p=0.942>.05$), indicating that the EG did not perform significantly better than the CG regarding the measure of the writing fluency.

Table 16. Paired samples t-test for fluency in the pre-and post-test for the EG and the CG

Measure	Group	Pre-test		Post-test		t	df	p (2-tailed)
		M	SD	M	SD			
W/M	EG	6.7937	1.2331	6.8008	0.7574	-0.038	35	0.97
	CG	6.5319	1.1774	6.7849	1.0459	-1.421	33	0.165

As seen in Table 16, the results revealed that students in the EG gained a certain degree of improvement in the W/M. However, the difference in writing fluency in the two tests was insignificant ($p=0.97>.05$). Likewise, the production of W/T increased slightly in the CG. However, no significant difference was detected between the two tests in the measure of fluency ($p=0.165>.05$). The implementation of DA yielded no significantly positive impact on students' writing fluency, indicating that students under the intervention of DA were unable to produce numerous words within the same allocated time.

4.3 Results of Interviews

After the intervention, the interviews were conducted to probe the students' perceptions of DA. The thematic analysis results yielded four themes: boosting writing motivation, change in writing attitude, improving learning efficiency, and suggestions to make up for limitations. The interview data provided substantiating evidence for the findings derived from the conducted tests.

Concerning the first theme, "boosting writing motivation," S1 claimed, "Because this model is combined with the teacher, the guidance from the online writing platform, and then coupled with peer assessment, our engagement and motivation will improve."

Echoing the precedent point, S2 pinpointed,

"Due to the feedback, I can notice these details and correct them in time. With the template essay, I can imitate or learn from other people's good expressions and some of the highlights, which can be applied to my writing."

They sensed that they were motivated to actively revise their writing due to multiple resources provided by the DA in the blended learning context.

Regarding the second theme, "change of writing attitude," S4 reported, "I feel that the DA in the blended learning mode is novel. I have not been exposed to it before, and then in this term, I like it more."

Corresponding with the previous statement, S2 mentioned,

"I prefer this model. After acquiring the teacher's feedback in the previous writing classes, we had to ask the teacher for details after class. This way is suitable for those who are not shy. I followed my peers' advice in this model and revised it, and then the score was instantly raised. I felt more and more confident."

They noted their preferences towards the intervention of DA in the blended learning context because they were prompted to participate more actively in writing activities and enjoyed the benefits, so their attitudes towards

writing have changed.

With respect to the third theme, "improving learning efficiency," S5 said,

"I prefer the dynamic assessment in the blended learning environment. Because I may not find any problems I encountered in the writing process, but if someone points them out, I can adjust my writing strategies or find ways to solve the problem without waiting a long time."

S6 confirmed a similar statement,

"I think it's conducive. By the pre-writing guidance, we will think about the topic from various perspectives. It can improve our critical thinking skills. Then, the teachers and peers gave you more feedback. It is efficient to integrate more sources of feedback. But it's probably less effective if it's just online or offline."

From their responses, this model contributes to applying their writing strategies in the compositions.

Regarding the fourth theme, "suggestions to make up for limitations," S4 responded to challenges and proposed powerful advice.

"Although we give assessments to others anonymously on the online platform, we sometimes may feel stressed to evaluate classmates' compositions. I suggest we evaluate the essays of other classes to reduce the pressure."

5. Discussion

The study examined to what extent DA impacts writing performance in the blended learning context and how participants perceived its effects. The findings confirmed the facilitative effects of DA on improving Chinese undergraduates' overall writing performance in the blended learning context. This finding echoes the results of some previous research (Alavi & Taghizadeh, 2014; Chen, 2021), which found that DA could contribute to a significantly positive effect on learners' writing performance. One possible explanation for the improvement is attributed to the mediation provided by integrated resources, which provided learners with various forms of meditation (Poehner, 2005). Each stage of learning and teaching is incorporated with the mediation by DA, and mediation is dynamic and interactive in that it is consistent with the student's responses to the mediation.

The study also explored the impacts of DA on writing CAF. The findings indicated no significant differences between the DA and traditional assessment of lexical diversity, syntactic complexity and fluency. However, it revealed that the DA boosted lexical complexity, especially lexical density, lexical sophistication, and accuracy.

These findings corroborate with studies by Chen (2021), reporting that the lexical density and sophistication in the learners' composition have significantly improved after DA embedded the blended writing courses. The results might be interpreted, as noted by Nabei and Swain (2002), given Vygotsky's sociocultural perspective, that both implicit and explicit feedback assisted learners more effectively perform in linguistic features. The intervention in the form of implicit and explicit feedback was beneficial in facilitating the student's internalization of advanced words.

However, no significant improvement was observed in the lexical diversity. Similar findings were also found in studies by Qian et al. (2023), revealing no significant positive effect of DA on students' writing performance regarding lexical diversity. The results are ascribed to Skehan's (1998) Limited Attentional Resources model. Due to their limited attentional resources, learners might not be able to pay attention to all dimensions of lexical complexity simultaneously. Instead, when the capacity of their attentional resources reaches its limits, they must decide which sub-dimensions to focus on. In the present study, lexical density and sophistication have witnessed significant increases. Students prioritized lexical density and sophistication over lexical diversity. In addition, learners at this level of language proficiency have probably reached a plateau in lexical word learning (Jullian, 2000). With a relatively higher level of English, students reached the plateau period of vocabulary acquisition earlier and tended to use the same words to lose fewer marks in exams, which, to a certain extent, led to the phenomenon that the vocabulary diversity of the compositions written was weakened.

As for syntactic complexity, despite the slight increase, the indices of synaptic complexity revealed no significant differences. The results demonstrated that the application of DA failed to impact syntactic complexity positively. The findings resemble those of Lin and Yang (2023), proposing that the effect of a combination of teachers' assessment and peer assessment on syntactic complexity falls short of significance. Prior research has verified that in second-language learning, lexical progress often precedes syntactic progress (Mazgutova & Kormos, 2015). It might also gain support from Van's (1991) point of view, positing that they could not be distributed evenly due to limitations of attentional resources, resulting in competition among subsystems. When students focus on the ability to use complex vocabulary, their tendency to produce more complicated syntactic structures is weakened. As underscored by Wu and Lei (2018), the development of syntactic complexity is

nonlinear in an unpredictable way, which takes a longer time to investigate. As a result, acquiring syntactic knowledge may take more time than acquiring vocabulary.

The findings of accuracy accord with that of Ebadi and Rahimi (2019), who reported a great extent of development of learners' writing accuracy after conducting DA sessions. The reason is based on ZPD. Learners may reveal emerging functions that they have not yet internalized under conditions of collaborative interaction. The students could have discerned linguistic structures that were situated within their ZPD and those extended beyond it. The gaps in the students' linguistic features might have been filled in properly and promptly.

A plausible explanation for the unfavored results of fluency is from the trade-off hypothesis proposed by Skehan (1988), who suggests a trade-off effect among different aspects of learners' written production. Writing accuracy and complexity require learners to perform a rule-based system for syntactic processing, while written fluency demands learners to extract existing language blocks in their memory-based system. Hence, as stated by Skehan (1988), to ensure written fluency, linguistic information should become long-term memory to extract effectively with minimal attentional resources whenever needed. Accordingly, attentional resources were predominantly devoted to the focus on language, thus improving complexity or accuracy. In this case, students overlooked the speed of their production and how to enhance it simultaneously with the accuracy and complexity of their production. Moreover, during the dynamic assessment writing training in the blended learning context, the automatic writing evaluation mainly involves the corrections of surface-level problems like language use rather than the speed of production.

The findings derived from the interviews revealed that most students perceived it as a facilitator contributing to developing their English proficiency, which correspond with the results of studies like those of Chen (2021) and Rezai et al. (2022), reporting that learners had positive attitudes and perceptions towards the efficacy of DA in promoting students' writing performance. Owing to the interaction, students can actively engage in the writing process, which boosts the internalization to address other issues and become capable of independent development, theoretically grounded in Vygotsky's sociocultural theory (1978). In this regard, students noticed the gaps between their current and higher levels and revised their compositions. Students could also observe how their writing abilities have improved and how ZPD has been expanded in writing. Students also demonstrated that their writing interest, confidence, and motivation had been enhanced. As Lidz and Gindis (2003) stated, students can complete tasks in shared or joint activities through mediated communication in DA. It creates a community for English writing in which teachers and students take on different roles, and students are encouraged to undertake as much responsibility as possible for writing so that both learners are motivated to engage in writing.

6. Conclusion

This study sets out to explore the effects of DA on Chinese undergraduates' writing performance consisting of overall writing performance and CAF and further investigate how students perceive the application of DA in the blended learning context. The findings suggested a statistically significant improvement in the students' overall writing performance. Also, it revealed that the application of DA facilitated lexical complexity, including lexical density, lexical sophistication, and accuracy, but failed to affect lexical diversity, syntactic complexity and fluency positively. Besides, the interview findings unveiled students' positive perceptions about the efficacy of DA. In a nutshell, it might offer a feasible approach for teaching and assessment in the blended learning context, and it would benefit students to enhance their English writing.

In light of these findings, this study offers implications for the instruction and assessment practice in the digital age. Firstly, incorporating the DA into blended learning in the writing mode allows students to receive various feedback resources. This mode can make the student-centered, process-oriented, and multiple assessments possible, which is a proficient way for English writing classes. As DA enables the teachers to pinpoint students' weaknesses and offer mediation when and where needed (Rezaee et al., 2019), teachers need to be sensitive to the features of the intervention, identify the ZAD and ZPD of students and provide intervention appropriately. Moreover, DA in the blended learning context could lessen their negative feelings regarding writing. Thus, it can create a peaceful and relaxing learning environment for students.

Although the study contributes to writing instruction, certain limitations need to be considered and addressed in future research. Firstly, the period is not long enough to get the expected results. Even though some valuable findings have been confirmed, the other results still need a much more extended period to testify. The period of future studies could be extended to examine how students' written English proficiency changes over time. Another limitation concerns the relatively small sample size included in the current study. Only 70 sophomores from two intact classes were involved in the study, and all of them were chosen from one normal university,

which might potentially impact the applicability of the findings. Future studies could expand the sample size appropriately to draw more approximate generalization conclusions. The sampling of participants with various majors from different universities in various regions should be considered.

Acknowledgments

The study has been supported by the Innovation Fund for Postgraduates of Jiangxi Normal University——Investigating the Effects of Online Peer Feedback on English Majors' Writing Performance from the Perspective of Dynamic Assessment (YJS2022073)

References

Alavi, S. M., & Taghizadeh, M. (2014). Dynamic Assessment of Writing: The Impact of Implicit/Explicit Mediations on L2 Learners' Internalization of Writing Skills and Strategies. *Educational Assessment, 19*(1), 1-16. https://doi.org/10.1080/10627197.2014.869446

Antón, M. (2009). Dynamic Assessment of Advanced Second Language Learners. *Foreign Language Annals, 42*(3), 576-598. https://doi.org/10.1111/j.1944-9720.2009.01030.x

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A Meta-Analysis of Three Types of Interaction Treatments in Distance Education. *Review of Educational Research, 79*(3), 1243-1289. https://doi.org/10.3102/0034654309333844

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101. https://doi.org/10.1191/1478088706qp063oa

Chen, D. D. (2021). The Impact of Online Peer Feedback on English Writing Performance from the Perspective of Dynamic Assessment. *Technology Enhanced Foreign Languages, 2*, 17-23, 3.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in Writing: Generating Text in L1 and L2. *Written Communication, 18*(1), 80-98. https://doi.org/10.1177/0741088301018001004

Dakduk, S., Santalla-Banderali, Z., & Van Der Woude, D. (2018). Acceptance of Blended Learning in Executive Education. *SAGE Open, 8*(3), 215824401880064. https://doi.org/10.1177/2158244018800647

Denscombe, M. (2007). *The Good Research Guide for Small-Scale Social Research*, (3rd ed.). Maidenhead: Open University Press.

Ebadi, S., & Rahimi, M. (2019). Mediating EFL learners' academic writing skills in online dynamic assessment using Google Docs. *Computer Assisted Language Learning, 32*(5-6), 527-555. https://doi.org/10.1080/09588221.2018.1527362

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity, and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*(1), 59-84. https://doi.org/10.1017/S0272263104026130

Hadidi, A. (2023). Comparing summative and dynamic assessments of L2 written argumentative discourse: Microgenetic validity evidence. *Assessing Writing, 55*, 100691. https://doi.org/10.1016/j.asw.2022.100691

Kushki, A., Nassaji, H., & Rahimi, M. (2022). Interventionist and interactionist dynamic assessment of argumentative writing in an EFL program. *System, 107*, 102800. https://doi.org/10.1016/j.system.2022.102800

Lantolf, J. P., & Poehner, M. E. (2011). Dynamic assessment in the classroom: Vygotskian praxis for second language development. *Language Teaching Research, 15*(1), 11-33. https://doi.org/10.1177/1362168810383328

Lidz, C. S., & Gindis, B. (2003). Dynamic assessment of the evolving cognitive functions in children. In A. Kozulin, B. Gindis, V. S. Ageyev, & S. M. Miller (Eds.), *Vygotsky's educational theory in cultural context*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511840975.007

Lin, L., & Yang, Y. X. (2023). The Effect of Teacher-student Collaborative Assessment on L2 Writing Performance. *Foreign Language Education, 44*(3), 72-78.

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu

Jullian, P. (2000). Creating word-meaning awareness [J]. *ELT Journal, 54*(1), 37-46. https://doi.org/10.1093/elt/54.1.37

Mackey, A., & Gass, S. M. (2016). *Second language research: Methodology and design*. New York: Routledge.

Mauludin, L. A. (2018). Dynamic assessment to improve students' summary writing skill in an ESP class. *Southern African Linguistics and Applied Language Studies, 36*(4), 355-364. https://doi.org/10.2989/16073614.2018.1548296

Mazgutova, D., & Kormos, J. (2015). Syntactic and lexical development in an intensive English for Academic Purposes programme. *Journal of Second Language Writing, 29*, 3-15. https://doi.org/10.1016/j.jslw.2015.06.004

Min, H. (2005). Training students to become successful peer reviewers. *System, 33*, 293-308. https://doi.org/10.1016/j.system.2004.11.003

Nabei, T., & Swain, M. (2002). Learner Awareness of Recasts in Classroom Interaction: A Case Study of an Adult EFL Student's Second Language Learning. *Language Awareness, 11*(1), 43-63. https://doi.org/10.1080/09658410208667045

Poehner, M.E. (2005). *Dynamic assessment of oral proficiency among advanced L2 learners of French*. (Unpublished doctoral dissertation). The Pennsylvania State University, University Park, PA.

Poehner, M. E. (2008). *Dynamic Assessment: A Vygotskian approach to understanding and promoting L2 development*. Berlin: Springer Science and Business Media. https://doi.org/10.1007/978-0-387-75775-9

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing, 26*, 10-27. https://doi.org/10.1016/j.jslw.2014.09.003

Polio, C., & Yoon, H. (2018). The reliability and validity of automated tools for examining variation in syntactic complexity across genres. *International Journal of Applied Linguistics, 28*(1), 165-188. https://doi.org/10.1111/ijal.12200

Qian, X., Meng, Y. R., & Yue, Z. (2023). Investigating the effectiveness of DA-embedded blended teaching on students' development of paraphrasing skill of academic English. *Foreign Language Education, 44*(02), 63-68.

Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics*. London: Routledge. https://doi.org/10.4324/9781315656762

Rahimi, M., Kushki, A., & Nassaji, H. (2015). Diagnostic and Developmental Potentials of Dynamic *Assessment* for L2 Writing. *Language and Sociocultural Theory, 2*(2), 185-208. https://doi.org/10.1558/lst.v2i2.25956

Rashidi, N., & Bahadori Nejad, Z. (2018). An Investigation into the Effect of Dynamic Assessment on the EFL Learners' Process Writing Development. *SAGE Open, 8*(2), 215824401878464. https://doi.org/10.1177/2158244018784643

Rezaee, A. A., Alavi, S. M., & Razzaghifard, P. (2019). The impact of mobile-based dynamic assessment on improving EFL oral accuracy. *Education and Information Technologies, 24*(5), 3091-3105. https://doi.org/10.1007/s10639-019-09915-1

Rezai, A., Naserpour, A., & Rahimi, S. (2022). Online peer-dynamic assessment: An approach to boosting Iranian high school students' writing skills: a mixed-methods study. *Interactive Learning Environments,* 1-19. https://doi.org/10.1080/10494820.2022.2086575

Shrestha, P., & Coffin, C. (2012). Dynamic assessment, tutor mediation and academic writing development. *Assessing Writing, 17*(1), 55-70. https://doi.org/10.1016/j.asw.2011.11.003

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press. https://doi.org/10.1177/003368829802900209

Tzuriel, D., & Shamir, A. (2002). The effects of mediation in computer-assisted dynamic assessment. *Journal of Computer Assisted Learning, 18*(1), 21-32. https://doi.org/10.1046/j.0266-4909.2001.00204.x

Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98*(1), 3-53. https://doi.org/10.1037/0033-295X.98.1.3

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing, 26*(3), 445-466. https://doi.org/10.1177/0265532209104670

Wu, X., & Lei, L. (2018). A meta-analysis of L2 writing proficiency and syntactic complexity. *Modern Foreign Languages, 41*(4), 481-492.

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing Class. *Journal of Second Language Writing, 15*(3), 179-200. https://doi.org/10.1016/j.jslw.2006.09.004

Yusoff, S., Yusoff, R., & Md Noh, N. H. (2017). Blended Learning Approach for Less Proficient Students. *SAGE Open, 7*(3), 215824401772305. https://doi.org/10.1177/2158244017723051

Zhang, Y. H. (2012). *Dynamic assessment in college English writing class*. (Unpublished doctoral dissertation). Shanghai International Studies University, Shanghai, China.

Appendix

Rubric Writing

Band	Score	Description
5	20-18	<u>Effective Communication Accuracies</u> The writing effectively addresses the writing tasks. It demonstrates a well-developed logical organizational structure with clearly stated main ideas and sufficient supporting details. It has almost no errors of vocabulary, spelling, punctuation or syntax, and it displays an adequate ability to use the language with appropriacy. No difficulty is experienced by the reader.
4	17-15	<u>Good Communication with Few Inaccuracies</u> The writing adequately addresses almost all of the writing tasks, though it deals with some parts more effectively than others. It demonstrates a generally well-developed logical organizational structure with main ideas and supporting details. It has relatively few significant errors of vocabulary, spelling, punctuation or syntax, and it displays an ability to use the language with appropriacy. Very little difficulty is experienced by the reader.
3	14-12	<u>Passable Communication with Some Inaccuracies</u> The writing adequately addresses most of the writing tasks. On the whole, it demonstrates an adequately developed organizational structure, though there may occasionally be a lack of relevance, clarity, consistency or support. It has occasional errors of vocabulary, spelling, punctuation or syntax, which may, from time to time, obscure meaning, and for the most part it displays some ability to use the language with appropriacy. Occasional difficulty is experienced by the reader.
2	11-9	<u>Problematic Communication with Frequent Inaccuracies</u> The writing only addresses some of the writing tasks. It demonstrates an inadequate organizational structure, and there may quite often be a lack of relevance, clarity, consistency or support. It has frequent errors of vocabulary, spelling, punctuation or syntax, and it displays a limited ability to use the language with appropriacy. Some difficulty is experienced by the reader.
1	8-6	<u>Almost No Communication</u> The writing almost completely fails to address the writing tasks. It has neither an organizational structure nor coherence. Almost all sentences contain errors of vocabulary, spelling, punctuation or syntax, and it displays no ability to use the language with appropriacy. Even after considerable effort on the part of the reader, the text is largely incomprehensible.