Language Complexity, Accuracy and Fluency in Different Types of Writing Paragraph: Do the Raters Notice Such Effect[*]

Alireza Ahmadi
Associate professor, Shiraz University

Sahar Alavi Zahed[**]
Phd in Elt, Shiraz University (corresponding author)

Abstract

The aim of the present study was to investigate the effects of two types of paragraph on EFL learners' written production. It addressed the issue of how three aspects of language production (i.e. complexity, accuracy, and fluency) vary among two types of paragraphs (i.e. paragraphs of chronology and cause-effect) written by EFL learners. Thirty intermediate level learners of English participated in the study. Each learner wrote the two specified types of paragraphs in the final exam of their writing course. In the first phase of the study, separate paired t-tests were conducted on each dependent variable to see whether there were any statistically significant differences in measures of complexity, accuracy, and fluency across the paragraph types. In the second phase of the study, to investigate if the raters detect the inconsistencies in the complexity, accuracy, and fluency of paragraphs written by a learner, four raters were asked to rate 12 paragraphs written by six learners whose paragraphs differed extremely in one of the three features. The findings revealed that EFL learners performed significantly better in paragraphs of chronology than the paragraphs of cause-effect in terms of fluency and accuracy. However, the analysis of complexity measures showed that there was no significant difference between the two types of paragraphs. In the qualitative analysis, it was found that raters did not consistently consider these three features in their examining the quality of the paragraphs. They paid attention to qualities such as coherence, cohesion and unity more consistently.

Key words: Types of paragraphs, EFL learners, Complexity, Accuracy, Fluency

Introduction

It is believed that second language performance could be explained through features of complexity, accuracy and fluency (CAF) (Ellis, 2003, 2008; Ellis & Barkhuizen, 2005; Freeman, 2009; Skehan, 1998). These three concepts have been used in investigating learners' language performance, both in oral and written forms.

As for the origin of the three concepts, In 1980s, a distinction was made between the fluency and accuracy of language use (Brumfit, 1984). Skehan (1989) added the third concept (complexity) to the triad. Complexity refers to the elaborateness and variety of the produced language (Ellis 2003). It is estimated by considering the number of subordinate clauses per clause (Wigglesworth, 1997) and the number of subordinate clauses per T-unit (Mehnert, 1998). Accuracy refers to the production of error-free language. It is estimated by considering the percentage of error-free clauses (Foster & Skehan, 1996; Skehan & Foster, 1999; Yuan & Ellis, 2003) and the percentage of correct use of target features (Crookes, 1989; Wigglesworth, 1997). Fluency refers to the speed of language processing and the general language proficiency. It is estimated by considering the mean length of utterances (Kormos & Dénes, 2004).

The present study was to investigate if the variation in the type of paragraph affects the quality of paragraphs written by EFL learners in terms of complexity, accuracy, and fluency. Two chosen paragraph types are chronology and cause-effect. So, the study is to investigate the probable variation in CAF features in chronology and cause-effect paragraphs.

The following research questions are addressed to this end:

1. Are there any varieties of CAF features in two types of paragraph (chronology and cause-effect)?

2. Do the raters notice CAF features in assigning scores to writing samples of chronology and cause-effect?

Review of Literature

There is measurable processing competence underlying language proficiency. A model for assessing the processing competence in tests is cited in Van Moere (2012). It assumes that the processing

competence is operationally defined as automaticity, assessed in terms of CAF features of the performance. The model consists of some levels related to the input, the psycholinguistic processing performed by the test taker and the output. The input is a prompt or question in a test. It is related to the manner in which the input is presented and its content. These two factors affect the item difficulty. The following level of the model relates to the internal processing of language, linguistic units. Processing production deals with conceptualizing the message and forming the lexical items within a syntactic structure. The aspects of production which are objectively measurable are complexity, accuracy, and fluency.

Measures of grammatical accuracy, complexity, vocabulary, pronunciation, and fluency guide in discriminating the proficiency levels of learners. They have independent contributions to the overall impression of the learners. A combination of these features determines the overall proficiency of the learner. These estimated measures in the learners' production vary according to their proficiency level. More problematic features can be observed in the productions of learners with lower proficiency level than those with higher proficiency level (Iwashita, Brown, Mcnamara & O'hagan, 2008). The learners with higher proficiency engage more constructively demonstrating a range of speech functions such as suggestions, agreement or disagreement, explanations, and challenges (Gan, 2010).

Once learners pay attention to one aspect of language production, some other dimensions are affected. Paying attention to each measure of CAF may result in various effects. When one pays attention to accuracy, he will present slower and less complex production; however, it enables them to speak confidently. When one deals with, he will produce novel structures, but more errors in production. When he pays attention to fluency, he will focus less on accuracy and complexity. Furthermore, the measures are supportive in the sense that development in any one of these dimensions of proficiency might depend on the development of another. As an example, the vocabulary growth will increase by the rapid increase in grammatical development (Freeman, 2009; Skehan, 2003).

The measures do not work for both individuals and groups as the basis for an SLA index of development. As an example, the average length of error-free T-units works well at the level of group, but not necessarily for individuals. When the progress of a group is charted, a steady improvement (an ascending line) over a period of time appears. However, this ascending line does not appear when the progress of individuals is observed. Therefore, the assumption of steady improvement in target language fluency, complexity and accuracy does not exist. It should be noted that from a complex systems theory, this variability is an important source of information about the underlying developmental process (Larsen-Freeman, 2006, 2009).

Some raise the question about the role of group rather than the individual work in the measures of fluency, accuracy and complexity in teaching contexts where formative assessments of second language writing ability are required. But, discourse analysis of the written products produced in collaborative tasks and pair works has shown that collaboration just affects accuracy but not fluency and complexity (Taylor & Wigglesworth, 2009).

Task effect

Concerning the effect of prompts on test performance, the studies have indicated significant differences in syntactic complexity and fluency of the test takers' production in different types of prompts. The prompts that elicited longer and more complex production encouraged test-takers to explain their family circumstances or speculate about their future. The Family prompt was an example of these categories. The prompts with more factual content elicited shorter, less complex turns. They required the test takers to present new topics (Leaper & Riazi, 2013). Tasks which were more familiar to the learners and whose structures were clear, such as presenting personal information, led to higher accuracy and fluency than complexity. Tasks that needed manipulation of information led to more complexity. Narrative tasks presented higher complexity but lower accuracy and fluency (Skehan, 2009). Learners performed better in the instruction task than in the argumentative task in terms of fluency and accuracy. However, complexity measures in argumentative essays were higher than those in instruction essays (Rezazadeh, et al., 2011).

Narrative structured task which is defined by such characteristics as clear time line, a script, a story with a beginning, middle, and end, and appealing to what is familiar in the writer's mind and the presence of a problem solution structure has a significant effect on the accuracy and fluency of the learners' writing performances, but not on the complexity. Performance in the structured tasks is more accurate than performance in the less structured tasks. So, the existence of structure in a narrative facilitates the production of accurate utterances. Furthermore, task structure can lead to the production of more fluent language. When learners plan their performance before their production of structured tasks, they will be able to produce more fluent language (Seyyedi, et. al., 2014).

The knowledge of the genres is important for writing a text. For example, for writing argumentative texts, learners should use the knowledge-transforming strategy; it needs higher planning demands, processes and cognitive effort; however, for writing narratives, learners should use easier knowledge-telling strategy. It needs lower planning demands, processes and cognitive effort. Furthermore, the type of genre selected affects the structure of the text indicated by the use of connectives. The chronological relations of events in narrative texts limit the need to use a large number of different connectives. However, there are more needs of use of connectives in the argumentative texts (Favart & Chanquoy, 2007).

Literature-response topics affect the oral production of EFL learners. Shiriyan and Nejadansari (2014) focused on measures of CAF concerning this issue. They provided the experimental group with some pieces of short literature reading texts with interesting content and with appropriate readability to motivate them to talk about their similar experiences. The results showed that literature-based activities improved the CAF features of EFL learners' oral production. They argued that the materials to which feeling and emotion are attached produce more vivid and long lasting memories. So, literature-based activities provide an authentic context in which learners can increase their grammatical and lexical knowledge. Literature-based activities lead to more fluent and accurate second language oral production.

Ismail, Samad, Eng. and Noordin (2012) investigated the effects of task complexity on the grammatical accuracy and syntactic complexity of second language production. Task reasoning demand (TRD) was a dimension of task complexity in Robinson's Cognition Hypothesis (2005). It was the extent to which a task required learners to reason, to explain causations and to give justifications. In a low complexity task, the learners had a general idea on the content of the paragraph. In a high complexity task, the learners might access their schemata, choose the relevant information, build on the knowledge through interaction with other learners, talk about their preferences and their emotions, argue and reason, justify their arguments, and provide cause-effect relationships. The results of their study indicated that more task reasoning demand resulted in greater syntactic complexity than less task reasoning demand. However, more task reasoning demand showed significantly less grammatical accuracy than less task reasoning demand. It was mentioned that accuracy was gained at the expense of complexity.

There are some differences in the quality of language production among native and non-native speakers. Native speakers, who have an implicit knowledge of their language systems, maintain their fluency regardless of the structure or complexity of the task they are dealing with. They plan their production easily. Thus, they have fewer problems concerning the fluency of their production. They can also focus on the accuracy of their production more. Non-native speakers have more problems in planning the structure of what they are to produce. Since the events in a narrative task are presented in a predictable way, the burden in their processing can be eased by presenting a narrative task to the learners. Regarding complexity, more task complexity results in more syntactic complexity for nonnative speakers in comparison with native speakers (Foster & Tavakoli, 2009).

Therefore, far little empirical research has been undertaken on the effect of types of paragraph on the measures of fluency, complexity and accuracy, and more specifically on whether and how raters notice such features in writing productions and base their sores on such features. The present study, therefore, has focused on this issue.

Method

This study employed a mixed methods design, which combines quantitative and qualitative approaches in a single study. It is used with the purpose of expanding the breadth and range of the study (Ary, Jocabs & Sorensen, 2016).

Participants

Thirty under-graduate students, 13 males and 17 females, studying in Hakim Sabzevari University, Iran, participated in this study. All of the students majored in English and were selected based on their availability. They were at the age range of 19 to 24. All of them had prior experience in writing academic English paragraphs. They had taken writing courses to enhance their English writing ability, and they had had formal training in English writing skills. The course of writing, which they had passed, oriented them to the use of writing conventions and standards including the general skills of writing as well as the basic structure of writing essays (the thesis statement, specific support, organization, unity, coherence and cohesion).

Materials

Sixty paragraphs, 30 paragraphs of cause-effect and 30 paragraphs of chronology, written by the participants made the data of this study. In fact, each participant produced two samples of writing: one a chronology paragraph and one a cause-effect paragraph.

Raters

Four raters participated in the study. They were Ph. D. candidates in Shiraz University, and had learned English in an EFL context. They are described in detail below.

- Rater 1 had taught General English in schools, English institutes and universities for eight years. However, he had taught writing for one year at intermediate levels. As for the rating experience, he was familiar with TOEFL or similar standard rating scales, although he did not use them in practice. Furthermore, he was familiar with some rating scales prepared by English institutes, and had used them for two years.

- Rater 2 had taught General English in institutes for six years. Furthermore, he had taught writing for three years at advanced

levels. As for the rating experience, he was familiar with TOEFL or similar standard rating scales. He had also used them for two years.

- Rater 3 had taught General English in English institutes and university for nine years. However, she had not taught writing specifically. As for the rating experience, she was not familiar with TOEFL or similar standard rating scales or even the rating scales prepared by English institutes.

- Rater 4 had taught General English in the schools, English institutes and universities for eight years. Furthermore, she had taught writing for two years at advanced levels. As for the rating experience, she was familiar with TOEFL or similar standard rating scales for four years, and had used them for two years.

To summarize, the raters had more or less similar background. All of them had the same level of education, being Ph. D. candidates, and were experienced in teaching English though they differed to some extent in their rating experiences. However, they became familiar with TOEFL iBT rating scale before conducting the study and rating.

Rating Scale
TOEFL iBT holistic writing scale was used as a writing rubric for rating the learners' written paragraphs. It provides scores ranging from 0 to 5. It considers if the writing addresses the topic effectively; if it is well-organized and well-developed; if it uses clearly appropriate explanations, exemplifications and details; if it displays unity, progression and coherence; if it displays consistent facility in the use of language; and if it demonstrates syntactic variety, appropriate word choice and idiomaticity, having minor grammatical and lexical errors.

Data Collection Procedure
Sixty samples were gathered from the participants in their final exam session of their Writing Course. To test their ability in writing English paragraphs, they were asked to write two types of paragraphs, namely paragraphs of chronology and paragraphs of cause-effect. As for the paragraph of chronology, the students were supposed to write on a sufficiently interesting, exciting or unusual experience they had in their high school or university. For the paragraph of cause-effect, they were supposed to give reasons for their agreement or disagreement on the

claim that "the era of silver screen is coming to an end and people will eventually lose interest in going to the cinema".

Data Analysis Procedure

In the first phase of the study, CAF measures were used to investigate the quality of the paragraphs. Schneider and Connor (1991) defined T-units as any independent clause and all its required modifiers, or any non-independent clause punctuated as a sentence (as indicated by end punctuation), or any imperative. Following Wigglesworth and Storch (2009), fluency was measured in terms of the average number of words, T-units and clauses per text.

To investigate the complexity, first following Foster and Skehan (1996), the proportion of clauses to T-units was estimated. Second, following Wolfe-Quintero, et. al. (1998), the proportion of dependent clauses to clauses (DC/C), indicating the degree of embedding in the text, was estimated.

To investigate the accuracy, the proportion of error-free T-units to all T-units (EFT/T) and the proportion of error-free clauses of all clauses (EFC/C) were estimated (Wigglesworth & Storch, 2009). In the present study, syntactical errors (e.g., errors in word order, missing elements) and morphological errors (e.g., verb tense, subject-verb agreement, errors in the use of articles and prepositions, and errors in word forms) were considered as error. Errors concerning the word choice were taken into account when the word used obscured the meaning. Errors in spelling and punctuation were ignored.

To conduct the intra-rater reliability, the researcher randomly selected 15 paragraphs of chronology and 15 paragraphs of cause-effect. As the estimates of the number of words, the T-units, the dependent clauses and overall clauses per text are central in estimating CAF measures, she investigated them again after two weeks' time span. The reliability coefficients estimated separately for each measure turned out to be .98, .96, .97 and .96 respectively. Five samples were randomly selected from each type of paragraphs. As for inter-rater reliability, a Ph. D. candidate was asked to investigate the previously mentioned concepts to estimate the inter-rater reliability. It was estimated as .98, .95, .97 and .95 respectively.

A series of paired t-test were conducted on all the measures. To control for Type 1 errors, a Bonferroni adjustment to the alpha level of .5 was used to judge the statistical significance. Furthermore, the effect size related to each t-test was reported.

In the second phase of the study, to investigate if the raters notice CAF features in rating writing samples, four raters were asked to rate 12 paragraphs written by six learners whose paragraphs differed extremely in one of the three features. In fact, after careful analysis of the sample writings produced by 30 participants, it became clear that six participants had produced interestingly different samples in terms of CAF features in chronology and cause-effect paragraphs. Therefore, the samples produced by these participants (12 samples) were selected for the second phase of the study. In addition to practicality, selection of such samples were mostly guided by the question of whether raters would notice such differences in CAF features in writing samples produced by the same individuals, and consequently estimate writing ability differently across the two types of paragraphs or not. In fact, the raters were asked to comment on their rating explaining why they had assigned a certain score to a writing sample. In other words, they were expected to talk about the criterion they considered in rating.

Results

Fluency measures

Table 1 shows the features of the paragraphs. It indicates that the learners have used a larger number of words, T-units and clauses in paragraphs of chronology compared to the paragraphs of cause-effect.

Table 1. Fluency Measures

Fluency	Type of paragraph	N	Mean	Std. Deviation	T test	Sig	eta squared
Average words per text	Chronology	30	142.8	35.74	1.99	.05	.12
	Cause-effect	30	130.36	28.87			
Average T-units per text	Chronology	30	13.03	4.05	5.35	.000	.5
	Cause-effect	30	8.93	2.62			
Average clauses per text	Chronology	30	16.41	4.50	5.31	.000	.49
	Cause-effect	30	11.51	3.13			

However, to see whether such differences are significant, a number of paired sample t-tests were employed. The results of the paired sample t-tests shown in Table 1 reveal that the differences in the written paragraphs are statistically significant in three measures (words per text, T-units per text, and clauses per texts). To control for Type 1 errors, the estimated Bonferroni adjustment turned out to be .0125. Based on this correction, when the results in Table 1 are examined more carefully, we will see that for the average words per text, the differences can not be safely considered as significant, though in other two cases the results are significant. Overall, fluency is significantly higher in the paragraph of chronology than that of cause-effect paragraph.

The effect size values were calculated and reported in the table. Given our eta squared value of .12, .5 and .49 and following Cohen's (1988) guideline for interpreting them, it can be concluded that there were moderate effects in the estimates.

Complexity measures
As it is evident in Table 2, there are more clauses per T-unit in paragraphs of chronology than the paragraphs of cause-effect. However, the dependent clauses percentage is higher in paragraphs of cause-effect.

The results of the paired sample t-tests shown in Table 2 reveal that the difference in the written paragraphs is not statistically significant in the two measures (clauses per T-unit, and dependent clauses percentage). The magnitude of the differences in the means related to dependent clauses percentage is very small (eta squared = .0004). Thus, it can be said that the participants' writing samples are the same in terms of complexity feature across the two types of paragraphs.

Table 2. Complexity Measures

Complexity	Type of paragraph	N	Mean	Std. Deviation	T test	Sig	eta squared
Clauses per T-unit	Chronology	30	1.32	.21	1.083	.288	.04
	Cause-effect	30	1.28	.26			
Dependent Clauses Percentage	Chronology	30	22.31	11.57	-.112	.911	.0004
	Cause-effect	30	22.63	13.34			

Accuracy measures
As Table 3 shows, the mean scores of the paragraphs of chronology for both measures of accuracy are higher than the mean scores for paragraphs of cause-effect.

Table 3. Accuracy Measures

Accuracy	Types of paragraph	N	Mean	Std. Deviation	T test	Sig	eta squared
Error free T-units percentage	Chronology	30	62.94	19.19	2.522	.017	.18
	Cause-effect	30	52.87	21.58			
Error free clauses percentage	Chronology	30	70.16	16.54	2.338	.026	.16
	Cause-effect	30	61.79	19.06			

The results of the paired t-test for complexity variables (Table 3) show that the two paragraph types differ with regard to error free T-units percentage and error free clauses percentage. Even when Bonferroni adjustment is applied (.02 in this case), the eta squared statistics of.18 and .16 indicate large effect sizes in estimates of error free T units percentage and error free clauses percentage, respectively. Thus, the results for accuracy indicate that paragraphs of chronology are more accurate than paragraphs of cause-effect.

Table 4 shows the CAF measures in paragraphs of chronology and cause-effect written by six EFL learners. Moreover, regarding the

paragraphs written by participant A, Rater 1 believed that the cause-effect paragraph had higher quality than that of chronology paragraph. He focused on the accuracy, organization, cohesion and coherence of the paragraphs. Rater 2 believed that both paragraphs of cause-effect and chronology suffered from the lack of grammatical, developmental and organizational features and detailed explanations to the same extent. Rater 3 believed that the participant's paragraph of chronology had higher quality than the paragraph of cause-effect. According to her, both paragraphs were short; they displayed unity, organization, and complex syntactic structures. However, the errors in paragraph of chronology did not lead to obscurity. Furthermore, paragraph of cause-effect needed more explanation. Rater 4 suggested that both types of paragraph suffer from insufficient exemplifications, details and organization.

Regarding the paragraphs written by participant B, rater 1 believed that the cause-effect paragraph had higher quality than that of chronology. It was more organized and enjoyed higher readability, cohesion and coherence. According to rater 2, cause-effect paragraph had higher quality than that of chronology. The paragraph of chronology was not accurate and fluent. Furthermore, it contained non-related information. However, the paragraph of cause-effect was well-organized, well-explained, well-progressed and intelligible. Rater 3 believed that paragraph of chronology had higher quality than that of cause-effect. Although showing good unity and coherence, the paragraph of cause-effect had poor grammatical structures and inappropriate choice of words. Rater 4 considered both paragraphs equivalent in the length, grammatical and lexical errors.

Table 4. Fluency, complexity and accuracy features in paragraphs of chronology and cause-effect written by six EFL learners

Writing features	Type of paragraph	Participant A	B	C	D	E	F
Fluency							
Average words per text	Chronology	79	154	124	158	208	188
	Cause-effect	91	137	157	153	149	118
Average T-units per text	Chronology	7	13	12	18	15	19
	Cause-effect	7	8	11	8	12	10
Average clauses per text	Chronology	9	15	14	18	17	22
	Cause-effect	8	8	11	14	17	12
Complexity							
Clauses per T-unit	Chronology	1.28	1.15	1.66	1	1.41	1.15
	Cause-effect	1.14	1	1	.75	1.54	1.2
Dependent Clauses Percentage	Chronology	22.22	13.33	14.28	0	29.41	13.63
	Cause-effect	12.5	25	0	42.85	35.29	16.66
Accuracy							
Error free T-units percentage	Chronology	71.42	76.92	50	77.77	53.33	57.89
	Cause-effect	0	37.5	72.72	62.5	66.66	60
Error free clauses percentage	Chronology	77.77	80	57.14	77.77	58.82	59.09
	Cause-effect	12.5	37.5	72.72	78.57	76.47	66.66

Concerning the paragraphs written by participant C, rater 1 believed that the score of cause-effect paragraph was higher than that of chronology. Although both paragraphs had the same level of accuracy and organization, that of cause-effect had more cohesion, coherence and readability. Rater 2 believed that the paragraph of chronology had higher quality than that of cause-effect. It had more coherence and cohesion. Furthermore, the idea was more explained. Rater 3 believed that the score of paragraph of cause-effect was higher than that of

paragraph of chronology. It had more coherence, unity and complex structures. Rater 4 suggested that the paragraph of cause-effect was higher than that of chronology. It displayed more unity and coherence.

As for the paragraphs written by participant D, rater 1 believed that the score of cause-effect paragraph was higher than that of chronology. Although both paragraphs had the same level of accuracy, readability, cohesion and coherence, the paragraph of cause-effect was more organized. According to rater 2, both paragraphs had equal scores. They were equally fluent and accurate. They had coherence, cohesion and appropriate explanations. Rater 3 believed that the paragraph of cause-effect had higher quality than that of chronology. Although the paragraph of chronology was well organized, coherent and complex syntactically, it needed more details and more elaboration. Furthermore, the paragraph of cause-effect had more cohesion, supporting ideas, coherence, unity and organization. It contained less grammatical errors. According to rater 4, both paragraphs had equal scores. They were well organized. They displayed unity, coherence and cohesion.

Concerning the paragraphs written by participant E, rater 1 believed that the score of cause-effect paragraph was higher than that of chronology. The paragraph of cause-effect was more organized and had more accuracy, readability, cohesion and coherence. Although it was shorter, it followed the mechanics of writing more. According to rater 2, both paragraphs had equal scores. They contained errors. Furthermore, they lacked coherence, cohesion and sufficient explanations. Rater 3 believed that the paragraph of cause-effect had higher quality than that of chronology. The paragraph of chronology lacked cohesion, appropriate organization and intelligibility. Although it was long, it contained redundant words and structures and ambiguity. The paragraph of cause-effect contained inappropriate words, grammatical structures and run on sentences. However, it was well-organized. According to rater 4, both paragraphs had equal scores. They consisted of limited topic development, inadequate organization and errors.

As for the paragraphs written by participant F, rater 1 believed that the score of cause-effect paragraph was higher than that of chronology.

The paragraph of cause-effect was more accurate and readable. According to rater 2, both paragraphs had equal properties. They lacked well organization, accuracy, detailed explanations. Rater 3 suggested that the score of paragraph of cause-effect was higher than that of paragraph of chronology. The paragraph of chronology provided explanation and exemplification. However, it lacked coherence, grammatical and lexical accuracy. According to her, although paragraph of cause-effect contained obscure syntactic structures and was less-organized, it was more coherent, and contained more explanations. Rater 4 believed that the score of paragraph of chronology was higher than that of cause-effect. It was more organized. Furthermore, it had more unity, coherence and cohesion.

Discussion

The present study aimed at investigating varieties of CAF features in two types of paragraph (chronology and cause-effect). It further aimed at seeing whether raters notice such features in assigning scores to written products. The findings of the statistical analysis suggested that the choice of type of paragraph affects the measures of fluency and accuracy of writing. The findings of the study do not support what Freeman (2009) and Skehan (2003) believe in terms of limited attention resources. They believe that one cannot attend to all aspects of language such as CAF features at the same time. Thus, there is a trade-off between these different aspects. In other words, if one pays attention to the fluency, he cannot dedicate the same level of attention to the accuracy or complexity as well. However, in the present study learners could pay attention to both accuracy of form and fluency in paragraphs of chronology. This discrepancy may be attributed to the effect of the type of paragraphs in the written performance of the learners. It might also be related to the effect of the organization of paragraphs, since the two types of paragraphs were of different types of paragraph organization. Paragraphs of chronology had time organization and paragraphs of cause-effect had listing organization. Ghabanchi and Alavi (2011) pointed to the effect of paragraph organization on the written productions.

The result of the study is consistent with those of Shiriyan and Nejadansari (2014) and Seyyedi, et. al. (2014) who showed that

literature-based topics lead to more fluent and accurate second language oral production. However, such activities did not affect the text complexity. They argued that as these activities deal with emotion and feeling, learners have fewer problems in comparison to other activities. And as Skehan (2009) proposed tasks which are more familiar to the learners and whose structures are clear, such as presenting personal information, lead to higher accuracy and fluency than complexity. In this study, similar results were obtained. Learners outperformed in paragraphs of chronology in terms of fluency and accuracy. The paragraphs of cause-effect present less fluency, since they deal with factual content and require the test takers to present new topics (Leaper & Riazi, 2013).

The results can also be related to the concepts of task reasoning demand mentioned in Ismail, et. al. (2012) and planning demands mentioned in Favart and Chanquoy (2007). Paragraphs of cause-effect require learners to reason, explain causations and give justifications. So, they present more task reasoning and planning demands in comparison to paragraphs of chronology. Learners should access their schemata, choose the relevant information, build on the knowledge through interaction with other learners, talk about their preferences and their emotions, argue and reason, justify their arguments, and provide cause-effect relationships. So, these tasks show less grammatical accuracy.

The qualitative phase of the study focused on whether the raters notice such features in the writing samples while rating or not. The two paragraphs produced by the participant A had some noticeable differences in terms of the three features. While the cause-effect paragraph was a bit longer and enjoyed less complexity in terms of ratio of dependent clauses, the basic difference between the two types of paragraphs referred to the accuracy feature. The chronology paragraph was produced with much higher levels of accuracy as indicated in Table 4. Among the four raters, three of them (rater 1, 2, 3) considered the feature of accuracy in their ratings. Rater 4 totally ignored it. Although three raters considered accuracy in their ratings, just one rater (rater 3) correctly considered it as a distinctive factor in assessing paragraphs of chronology and cause-effect. Rater 3 considered not only the errors, but also the intelligibility of the erroneous structures, and assigned higher

score to the chronology paragraph because of higher accuracy. However, raters 2 and 4 assigned the same scores to both paragraphs and considered them to be the same in terms of different features. More interesting is the fact that the first rater thought that the paragraph of cause-effect was even more accurate (exactly the opposite of what the features in Table 4 indicate).

Concerning the paragraphs written by the participant B, Table 4 indicates that the chronology paragraph enjoys higher quality in all the three features and more particularly in the accuracy feature. However, only rater 3 partially noticed this difference and assigned a higher score to this paragraph. This rater mentioned that the cause-effect paragraph enjoyed less complexity, and failed to focus on differences in terms of the other features especially accuracy which is noticeably different in the two paragraphs. The other three raters were totally off-tarck and failed to notice the differences depicted between the two paragraphs. Rater 4 considered the two paragraphs to be similar, and Raters 1 and 2 considered the cause-effect paragraph to be better.

With regard to the paragraphs written by the participant C, Table 4 indicates that while the cause-effect paragraph was a bit longer in terms of the average words per text and enjoyed more accuracy, the basic difference between the two types of paragraphs referred to the complexity feature. The chronology paragraph was produced with much higher levels of complexity as indicated in Table 4. Three of the raters (rater 1, 3, 4) considered the quality of cause-effect paragraph as higher than that of chronology. Just one rater (rater 3) considered complexity as a differentiating feature in two paragraphs. However, she thought that the paragraph of cause-effect was more complex (exactly the opposite of what the features in Table 4 indicate). The other three raters totally ignored this feature.

Considering the paragraphs written by the participant D, Table 4 indicates that while the chronology paragraph was a bit longer and enjoyed a bit more accuracy in terms of error free T-units percentage, the basic difference between the two types of paragraphs referred to the complexity feature. The cause-effect paragraph was produced with much higher levels of complexity in terms of dependent clauses percentage as indicated in Table 4. Two of the raters (rater 1, 3) believed

that the paragraph of cause effect was of higher quality than that of chronology. Raters 2 and 4 assigned the same scores to both paragraphs and considered them to be the same in terms of different features. Just rater 3 considered complexity feature in her decision making. However, she thought that paragraph of chronology was more complex (exactly the opposite of what the features in Table 4 indicate).

Concerning the paragraphs written by the participant E, Table 4 indicates that while the cause-effect paragraph was a bit more complex and a bit more accurate, the basic difference between the two types of paragraphs referred to the fluency feature. The chronology paragraph was produced with much higher levels of fluency as indicated in Table 4. Two of the raters (rater 1, 3) believed that the paragraph of cause effect was of higher quality than that of chronology. Raters 2 and 4 considered the two paragraphs to be similar. Although three raters (rater 1, 2, 3) considered fluency in their ratings, just two raters (rater 1, 3) correctly considered it as a distinctive factor in assessing paragraphs of chronology and cause-effect.

With respect to the paragraphs written by the participant F, Table 4 indicates that while the cause-effect paragraph was a bit more complex and a bit more accurate, the basic difference between the two types of paragraphs referred to the fluency feature. The chronology paragraph was produced with much higher levels of fluency as indicated in Table 4. There was no consensus among raters regarding the overall quality of the paragraphs. Rater 1 and 3 considered the quality of paragraph of cause-effect as higher in comparison to that of chronology. Rater 2 believed that two paragraphs were similar. Finally, rater 4 believed that the paragraph of chronology had higher quality. As for considering fluency as a feature in two paragraphs, raters 2 and 3 did it. However, rater 2 didn't consider it as a distinctive feature.

To summarize, some raters did not notice CAF features in assessing the paragraphs. They considered two paragraphs as similar. Although some raters considered these feature as distinctive in assessing paragraphs, they considered them exactly the opposite of the actual features the paragraphs. Only, a limited number of raters noticed the features partially.

Conclusion and Implications

The purpose of the present study was to provide an insight into the effects of the types of paragraphs (chronology and cause-effect) on the CAF measures of EFL learners' written performance. The findings indicated that the participants performed significantly better in paragraphs of chronology than paragraphs of cause-effect in terms of fluency and accuracy. However, the analysis of complexity measures revealed that there was no difference between the two types of paragraphs, and any probable difference was the function of chance alone. It can be concluded that the choice of the type of paragraphs affects the quality of learners' performance in terms of complexity, accuracy, and fluency.

In the qualitative phase of the study, almost all of the raters considered cohesion, coherence, organization and unity in rating and assessing the quality of the paragraphs. However, CAF features of the paragraphs were ignored by some raters. Furthermore, among the raters who considered the CAF features of the paragraphs, some of them were not consistent in considering the features in their decision making on the quality of the paragraphs. As an example, rater 1 considered accuracy in examining the paragraphs of participant A, but not those of participant B. Similarly, rater 4 considered accuracy as a factor in evaluating paragraphs of participant B, but not those of participant A. Furthermore, among the raters who considered each of the features of complexity, accuracy, and fluency, some did it incorrectly (their mentioned features were the opposite of the actual features of the paragraphs). In sum, it might be concluded that raters consider more general features (cohesion, coherence and unity) in their ratings, and they don't pay sufficient attention to features of complexity, accuracy, and fluency.

In assessing the paragraphs, only one rater (rater 3) considered CAF features in her ratings, although among her ratings, two were the opposite of the actual qualities of the paragraphs. It is interesting to know that this rater had not taught writing, and was not familiar with TOEFL or similar standard rating scales. It seems that more experienced raters (in comparison to less experienced ones) consider general features rather than CAF in their ratings.

The study has some implications for material developers, teacher educators and teachers. Material developers should pay attention to the difference in the emerging patterns of CAF features across different types of paragraphs. They can devote supplementary parts and activities in the books for introducing the paragraph types in which learners show lower patterns of CAF features in students' performances. Teacher educators should make teachers aware of the emerging difference in the CAF features of students' performances across different types of paragraphs. In writing courses, teachers can plan to devote the sufficient amount of time required to present different types of paragraphs.

There are some suggestions for further studies. First, this study is conduced considering just one level of proficiency (intermediate). Other studies can investigate the CAF features in different types of paragraphs across learners with different proficiency levels. Second, one study can compare the performance of native speakers and non-native speakers in language performance in terms of complexity, accuracy, and fluency. Third, one study can investigate the effect of planning on the CAF features of different types of paragraphs. That is, one can investigate whether planning makes any difference in the estimating the three concepts in different types of paragraphs. Fourth, one study can investigate the features that raters consider while rating the paragraphs.

References

Ary, D., Jocabs, L., C., & Sorensen, C., K. (2016). *Introduction to research in education*. New York: Cengage Learning.

Brumfit, C. J. (1984). *Communicative methodology in language teaching*. Cambridge: Cambridge University Press.

Cohen, J.W. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.

Crookes, G. (1989). Planning and interlanguage variability. *Studies in Second Language Acquisition, 11*, 367-383.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford: Oxford University Press.

Ellis, R. (2008). *The study of second language acquisition (2nd edition)*. Oxford: Oxford University Press.

Ellis, R. and G. Barkhuizen. (2005). *Analysing learner language.* Oxford: Oxford University Press.

Favart, M., & Chanquoy, L. (2007). Cohesion devices as crucial tools of text composition: A comparison between 5th graders and expert adults. *Langue Française, 155*(3), 51-58.

Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, *18*(3), 299-323.

Foster, P. & Tavakoli, P. (2009). Native speakers and task performance: comparing effects on complexity, fluency, and lexical diversity. *Language Learning*, *59*(4), 866–896.

Gan, Z. (2010). Interaction in group oral assessment: A case study of higher- and lower-scoring students. *Language Testing*, *27*(4), 585–602.

Iwashita, N., Brown, A., Mcnamara, T.& O'hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24–49.

Ismail, L., Samad, A., Eng. W. & Noordin, N. (2012). Effects of task reasoning demand and task condition on learner written output in ESL classrooms. *International Journal of Education, 4*(4), 157-175.

Ghapanchi, Z & Alavi Z. S. (2011). The relation between paragraph organization and the topic progression used in English paragraphs selected

from native books on teaching writing. *Iranian EFL Journal, 7*(3), *243-253.*

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*, 145-164.

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, *27*(4), 590–619.

Larsen-Freeman, D. (2009). Adjusting Expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 579–589.

Leaper, D. & Riazi, M. (2014).The influence of prompt on group oral tests. Language Testing, 31(2), 1–28.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, *20*, 83-108.

Rezazadeh, M., Tavakoli, M. & Eslami, A. (2011). The role of task type in foreign language written production: Focusing on fluency, complexity, and accuracy. *International Education Studies, 4*(2), 169-176.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL, 43*, 1-32.

Seyyedi, K., Ismail, S., & Mohamed, A. (2014). The Effect of Task Structure on Second Language Learner's Narrative Writing Performance. *Journal of Foreign Languages, Cultures and Civilizations*, *2*(1), 41-53.

Schneider, M., and Connor, U. (1991). Analyzing topical structure in ESL essays: Not all topics are equal. *Studies in Second Language Acquisition*, *12*, 411-427.

Shiriyan, Z. & Nejadansari, D. (2014). The effect of literature-response activities on the complexity, accuracy, and fluency of Iranian EFL learners' L2 oral productions. *Journal of Applied Linguistics and Language Research, 1*(2), pp. 12-26.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skehan, P. (2009). Modeling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning*, *49*, 93-120.

Taylor, L. & Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts? *Language Testing, 26* (3), 325–339.

Van Moere, A. (2012).A psycholinguistic approach to oral language assessment. *Language Testing*, *29*(3), 325–344.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, *14*, 167-197.

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

Wolfe-Quintero, K., Inagaki, S., & Kim, H. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu, Hawaii: University of Hawaii Press.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity, and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*, 1-27.