

Effects of Task Types on the Writing Performance of Japanese English-as-a- Foreign-Language Learners across Proficiency Levels in terms of Complexity, Accuracy, and Fluency

FUJIWARA Yumi (u_fujiwara@yahoo.co.jp)*
Kyoto University of Education, Kyoto, Japan

Received: September 17, 2015, Accepted: December 20, 2015

Abstract

The purpose of this study was to investigate whether different types of second language (L2) writing tasks would differently affect the linguistic performance of Japanese EFL (English as a Foreign Language) university students across different proficiency levels. Subjects consisted of 30 Japanese EFL learners, who had been divided into two different language proficiency levels, CEFR level B1 ($n = 15$) and B2 ($n = 15$), on the Oxford Quick Placement Test. They were asked to perform individual writing tasks: i.e. narrative and argumentative task. In order to cancel the effect of task order, the participants were randomly divided into two groups with crossover task performance: viz., one group performed the narrative task the first week and the argumentative task the next, while the other group did the reverse. Then, the 60 completed texts were collected and measured for complexity, accuracy and fluency (collectively known as CAF). The method of statistical analysis consisted of a two-way-mixed-design ANOVA test performed over each CAF variable. The results from ANOVAs demonstrated that the argumentative task improved student performance significantly better than the narrative task for both proficiency groups with one exception: lower-proficiency learners (the B1 group) showed no significant difference in performance on the two tasks in the aspect of syntactic complexity. Additionally, the findings revealed that the B2 learners performed significantly better than the B1 learners, especially in terms of accuracy, on both tasks.

Keywords: writing performance, task type, proficiency level, complexity, accuracy, fluency

1. Introduction

The globalization of culture such as travel, business, and entertainment has cemented the position of English as the modern *lingua franca*. One outcome of this in the field of second language acquisition (SLA) research education is that since the 1980s, task-based language teaching (TBLT) has increasingly become more popular. In TBLT, a task is defined as “an activity which requires learners to use language, with emphasis on meaning, to attain an objective.”¹ In short, TBLT can give EFL (English as a Foreign Language) learners the opportunity to engage in practical, functional, and learner-centered activities that involve performing various kinds of tasks; EFL learners are prompted to use the knowledge and skills they have already acquired in the course of completing a task. More importantly, in active learning, they are given opportunities to explicitly notice the gap between what they want to say and what they can actually say in the target language.² That is to say, TBLT triggers more meaningful observation and reflection than traditional teaching methods wherein students tend to be passive participants. Thus, through active engagement in the right communicative task, EFL learners can enhance not only

their output but also their cognitive awareness. Writing activities *per se* require learners to express thoughts in their own words, leading them to observe their language use. Therefore, one would expect the task-based method to be especially well-suited to teaching writing to EFL learners, as well as to serving as an instructional tool for assessing writing performance.

However, implementing TBLT has not yet become fully practical in Japanese EFL classrooms, wherein the form-focused approach is still common.³ The conventional method of writing instruction involves teaching the translation of individual sentences.⁴ This can lead to students not receiving enough opportunities to practice written expression in their second language (L2). In addition, the exam-based education system of Japan places too much emphasis on grammatical accuracy as opposed to creative fluency, which means that Japanese EFL learners tend to deal with writing activities cautiously, and produce shorter texts with fewer errors instead of longer ones that might contain more errors.⁵ This limited quantity of writing is not enough for teachers to provide students with sufficient or adequate feedback. As a consequence, Japanese EFL students have a hard

time taking notice of what they need to learn: they end up making little progress. Thus, in order to help teachers to promote taking greater notice by students, and elicit sufficient output from students within the limited classroom time, we recommend the effective incorporation of TBLT into Japanese EFL classrooms, and the utilization of various types of communicative tasks in teaching writing.

The main focus of TBLT is the performance of tasks; in fact, there are many tasks conventionally used for this purpose in EFL settings. Given that task type is one of the most important factors affecting L2 writing performance,⁶ students levels of proficiency should also be taken into consideration when seeking the most appropriate task type to assign. To the best knowledge of the researcher, little research has tackled the issue of the relative effectiveness of different kinds of task at improving the linguistic aspects of Japanese EFL learners' writing performance according to their proficiencies. Accordingly, this study aims to explore the relationship between task type, L2 proficiency, and L2 performance, particularly focusing on the dimensions of complexity, accuracy, and fluency (CAF for short). CAF are three basic dimensions for describing L2 performance, proficiency, and development employed in several SLA studies.⁷ Generally speaking, it is natural to assume that as learners' language proficiency improves, their performance on writing tasks should also improve. Specifically, as learners become more proficient, their written work should become more syntactically and lexically complex, accurate, and fluent. For this purpose, this study employed two different widely-used task types: i.e. narrative and argumentative writing.

2 Methods

2.1 Purpose and main research question

This study was designed to examine the effect of each CAF dimension on Japanese EFL learners' writing performance across different proficiency levels. More precisely, the effect of task-type on each CAF variable for the results of two different writing tasks assigned to Japanese EFL university students with different levels of proficiency was elucidated. The statistical analysis used the software package R 3.1.0 for Windows, and

differences where $p < .05$ were considered significant.

2.2 Participants

In this study, participants were screened to have an at least intermediate level of proficiency on the grounds that they would thereby have acquired sufficient L2 ability to develop an idea and compose a creative essay, in English. The participants were chosen from 36 non-English major freshman and sophomore students at a Japanese national university (female: 21, male: 15; age-range: 19-23 years). All students had studied English as a foreign language for over six years as part of their formal Japanese education. None of them had studied English abroad for any period greater than a month. For the purpose of evaluating the general proficiency level of the participants, the Oxford Quick Placement Test (OQPT)⁸ was administered prior to their enrollment in the task phase. This test, developed by Oxford University Press and Cambridge ESOL, contains 60 questions on vocabulary, grammar and reading comprehension. After screening, 30 students were considered eligible for the study based on the following criteria: (i) eligible students were required to be able to follow the series of tasks or steps in order to complete the entire process; and (ii) their OQPT scores had to place them in either the B1 or B2 level of the CEFR (Common European Framework of Reference for Languages).⁹ CEFR defines six proficiency levels for a language user, from basic A1 to advanced C2. Broadly speaking, the B1 level is viewed as intermediate while B2 represents upper intermediate. Table 1 displays the independent t -test results for the two proficiency levels and OQPT scores for all participants as well as the mean (M), standard deviation (SD), maximum scores (Max) and minimum scores (Min), supporting the conclusion that the difference in OQPT scores was significant between these two groups ($t(28) = -8.82, p = .00, r = .86$).

2.3 Tasks and procedure

The participants were instructed to perform the two different writing tasks (narrative and argumentative) separately on two sequential weeks. The narrative task required them to tell a story using a six-panel illustration (Fig. 1) as a visual cue, whereas the argumentative task asked them to write an opinion essay providing reasons

Table 1: Results of t -test and descriptive statistics for OQPT scores

CEFR level	n	M	SD	$Max.$	$Min.$	t
B2	15	42.6	1.99	47	40	-8.82*
B1	15	34.93	2.71	39	30	
Total	30	38.77	4.55	47	30	

Note. * $p < .05$

and examples to support their argument. An illustration entitled: “Waiting for a bus” was chosen for the narrative cue because it was deemed to have a sufficiently complex storyline involving foreground/background information.¹⁰ The six panels recount a vivid story of the younger boys at first being prevented by a bunch of older boys from getting on a bus, and where their bus eventually was running more smoothly than one which the older boys got on. The topic for the argumentative writing task was a TOEFL iBT exam’s independent writing topic: “*People do many different things to stay healthy. What do you do for good health?*” The primary reason for selecting this topic was that it involved a familiar or everyday situation, requiring little specialized knowledge or vocabulary.

The study was conducted during a regular English class on two consecutive weeks. Students performed both tasks as part of their regular classroom instruction without being given special instructions, such as on the nature or purpose of the study, to avoid self-consciousness or hyper-awareness of the experiment. In addition, all participants were given the instructions in Japanese so as to avoid misinterpretation of what each task required. After the 30-minute OQPT placement exam prior to the talk, students were randomly divided into two equal groups, in order to allow randomization of task order. Subsequently, each group engaged in their assigned writing task without use of a dictionary or additional individual preparation time, and likewise, without prior

announcement, with the remaining task the following week. Each task was limited to 25 minutes, immediately after which their written texts were collected. On the week following the completion of both tasks, the participants received some corrective feedback, which included marking of common grammatical errors in their essays. In order to analyze quantitatively with regard to complexity, accuracy and fluency, the collected handwritten documents were converted into digital form: a total of 60 texts (i.e. two tasks for 30 participants) were transcribed and saved as text files.

2.4 Text measurements

There have been numerous previous studies which proposed different ways to measure linguistic features of language production in degrees of complexity, accuracy and fluency. The measurements employed to rate each dimension in this study are discussed in more detail below; however, each has been found to be among the most reliable ones for assessing L2 written samples.⁷ Identifying and counting production units (i.e. words, clauses and T-units) was the first step in the data analysis process. The reason was that compositions with greater numbers of words, clauses, and T-units are considered to be more fluent and complex, suggesting greater competence on the part of the writer. Therefore, this measure has been generally utilized in SLA research for the analysis of writing performance.

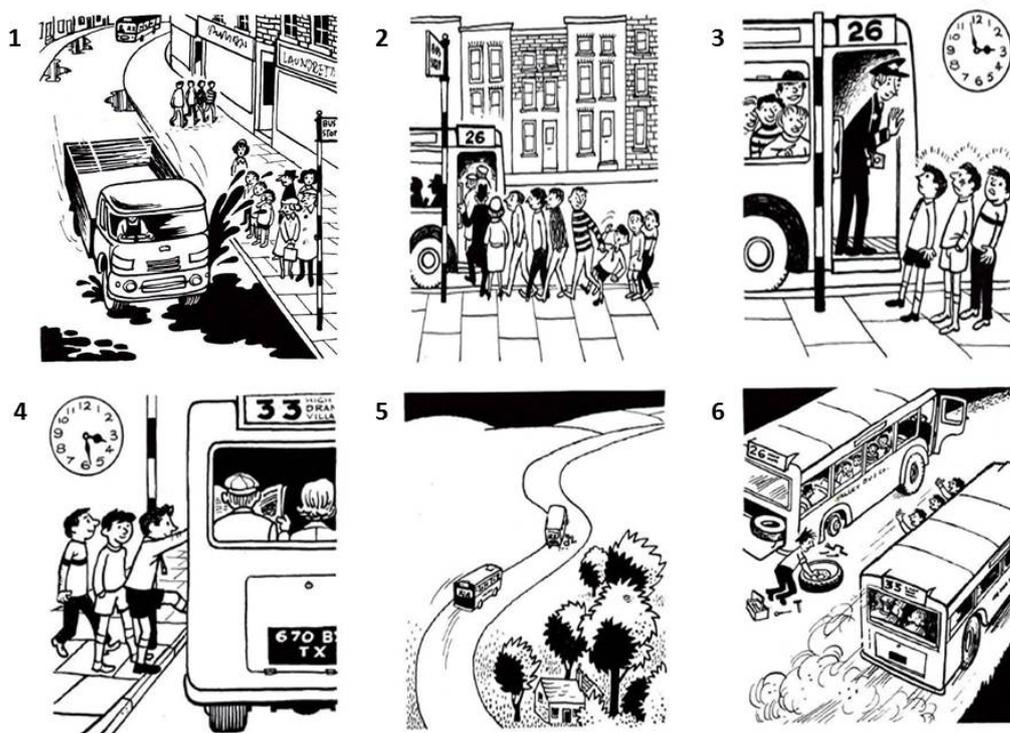


Fig. 1: A “Waiting for a bus” picture extracted from Heaton (1975: 54)

2.4.1 Coding for words, clauses and T-units

Vocabulary can be evaluated according to the total number of individual words (tokens), or as different forms of what is considered the same word (types). As for words, it is hard to tell how many words there are in a sentence. For example, the sentence: “*I forgot Mary but Mary did not forget me*” contains 9 tokens and 8 types because *Mary* occurs in it twice. Furthermore, whether *forget* and *forgot* should be counted as being of different type is debatable. Based on the concept of lemmatization, these words (i.e. *forget* and *forgot*) are classified as being of one type (i.e. *forget*), meaning that the sentence has 7 word types. Lemmatization is “a process wherein the inflectional and variant forms of a word lemma: their base form, or diary look-up form.”¹¹ In this study, word types (total unique words) were referred to as ‘lemmatized’ word types for the reason that “lemmatized text is an invaluable aid for semantic studies and others using analysis techniques involving repeating sequences of words or word pairs.”¹¹ For obtaining token and type information, each text was processed by a computer program called ‘v8an.’¹²

In addition, segmentation of each text into T-units/clauses was necessary to quantify L2 performance. A T-unit is defined as “a main clause plus all subordinate clauses and non-clausal structures attached or embedded in it.”¹³ Meanwhile, as for a clause, it is essential to identify the difference between independent and dependent clauses: certain coordinating conjunctions (e.g. *and*, *but*) serve to indicate the start of independent clauses, while dependent clauses begin with subordinating conjunctions (e.g. *because*, *although*). By definition,¹⁴ an independent clause indicates “a grammatical structure which contains a subject and a verb can stand on its own,” and a dependent includes “a finite or non-finite verb and at least one additional clause element of the following: subject, object, complement or adverbial.” Following Polio’s detailed guidelines (1997),¹⁵ T-unit/clause boundaries for every composition could be detected manually.

2.4.2 Measures to assess CAF (complexity, accuracy and fluency)

Regarding complexity, a further subdivision is generally made into syntactic and lexical complexity.¹⁶ This study also placed emphasis on both syntactic and lexical complexity in order to explore linguistic aspects of writing performance in more detail. Syntactic complexity refers to “the range of forms that surface in language production and the degree of sophistication of such forms,”¹⁷ indicating that its measure is based on subordination. For

addressing syntactic complexity, a widely adopted method was employed: viz., the ratio of clauses to T-units. Meanwhile, lexical complexity refers to “the degree of elaboration, the size, the breadth, width or richness of the L2 learner’s [lexical knowledge].”¹⁷ That is, complexity in vocabulary relates to variety in the words occurring in a given text. As a measure of lexical variation, the Type-Token ratio (TTR) is a widely recognized measure, whose validity is however frequently questioned due to its sensitivity to text length. Therefore, in this study, the Guiraud index¹⁸ was adopted as an alternative to TTR. The Guiraud index was devised for the purpose of reducing text-length dependency and is measured by dividing the number of types by the square root of the number of tokens (Types/ $\sqrt{\text{Tokens}}$). A higher index is meant to indicate greater lexical complexity.

Accuracy is referred to as “the ability to produce error-free speech.”¹⁷ Accuracy was measured in the way as described Ellis and Yuan (2004),¹⁹ using both general and specific measures: general accuracy was calculated based on the percentage of error-free clauses to all clauses, and specific accuracy based on the percentage of correctly used verbs. Error-free clauses mean clauses that have no errors in syntax, morphology, or lexical choice; however, it is quite difficult, other than in advanced learners, to detect error-free units in L2 performance. Therefore, only global errors were taken into consideration as errors for this analysis. Unlike local errors, global errors are ones which seriously affect the overall sentence organization, hindering the flow of communication (e.g. wrong word order, missing elements, awkward phrasing).²⁰ With regard to verb-related errors, this study basically utilized the error-tagging guidelines of the NICT JLE Corpus,²¹ which are based on oral transcripts of texts produced by more than 1,200 Japanese EFL learners. This error system provides a detailed error category for each part of speech. There are 11 categories of verb-related errors. Among these, a closer look into the participants’ individual writing led to consideration of only the relevant and high frequent error tags, with some modifications, a list of which is provided below (Table 2). The errors identified were classified manually by referring to the NICT JLE Corpus Error Tagging Guidelines.

Fluency, which is defined as “rapid production of language,”²² was measured in terms of words per T-unit. Taken together, the summary of measures to assess CAF is given below (Table 3). In addition, in order to ensure the reliability of results, the number of T-units, clauses and general/specific errors was counted twice by the same rater/researcher with a time interval of two months, respectively. Thus, simple correlation between the

number the rater counted was calculated as a measure of intra-rater reliability for the coding, yielding $r = .94, p < .05$ for T-units, $r = .91, p < .05$ for clauses, $r = .92, p < .05$ for general errors (i.e. global errors) and $r = .89, p < .05$ for specific errors (i.e. verb-related errors). The result revealed a positive correlation, and disagreements revealed from double-coding were then discussed with a second rater who is a Japanese EFL university instructor with over 20 years of teaching experience in Japan.

3. Data analysis and results

In order to answer the main research in this study, a two-way-mixed-design ANOVA test was conducted separately, with L2 proficiency (B1 and B2) as the between-subjects factor and task type (narrative and argumentative) as the within-subjects factor. The descriptive statistics for each dependent variable (Table 4), and boxplots of the distribution of each variable in both tasks across L2 proficiency (Fig. 2) were well demonstrated.

3.1 Complexity

Regarding syntactic/lexical complexity, the same pattern

Table 2: Verb-related error category

Error Category	Examples
Inflection	<i>I *spoke to the man.</i>
Subject-verb agreement	<i>There *is two people in the bus.</i>
Tense	<i>I *go to shopping two days ago.</i>
Form	<i>*Walk every day is good for the health.</i>
Lexical choice	<i>I will *do a speech tomorrow.</i>
Omission (Missing verbs)	<i>Without you, the plan would have *possible.</i>

Note. An asterisk (*) is used to label ungrammaticality.

Table 3: Variables used to assess CAF in the present study

Dependent	Measures
syntactic	the ratio of clauses to T-units
lexical complexity	the Guiraud index
accuracy (general)	the ratio of error-free clauses to clauses
accuracy (specific)	the percentage of correct verb forms
fluency	the number of words per T-unit

Table 4: Descriptive statistics concerning CAF variables

variable	level	n	Narrative Task		Argumentative Task	
			M	SD	M	SD
syntactic complexity	B2	15	1.34	0.17	1.67	0.42
	B1	15	1.30	0.13	1.42	0.22
	Total	30	1.32	0.15	1.54	0.35
lexical complexity	B2	15	5.91	0.60	6.65	0.74
	B1	15	5.93	0.56	6.19	0.95
	Total	30	5.92	0.57	6.42	0.87
accuracy (general)	B2	15	0.93	0.04	0.97	0.03
	B1	15	0.82	0.10	0.93	0.06
	Total	30	0.87	0.09	0.95	0.05
accuracy (specific)	B2	15	0.88	0.50	0.93	0.05
	B1	15	0.79	0.09	0.91	0.04
	Total	30	0.83	0.08	0.92	0.05
fluency	B2	15	8.56	1.62	10.94	2.44
	B1	15	8.17	1.03	9.32	0.98
	Total	30	8.36	1.35	10.13	2.00

Note. M = Mean, SD = Standard Deviation

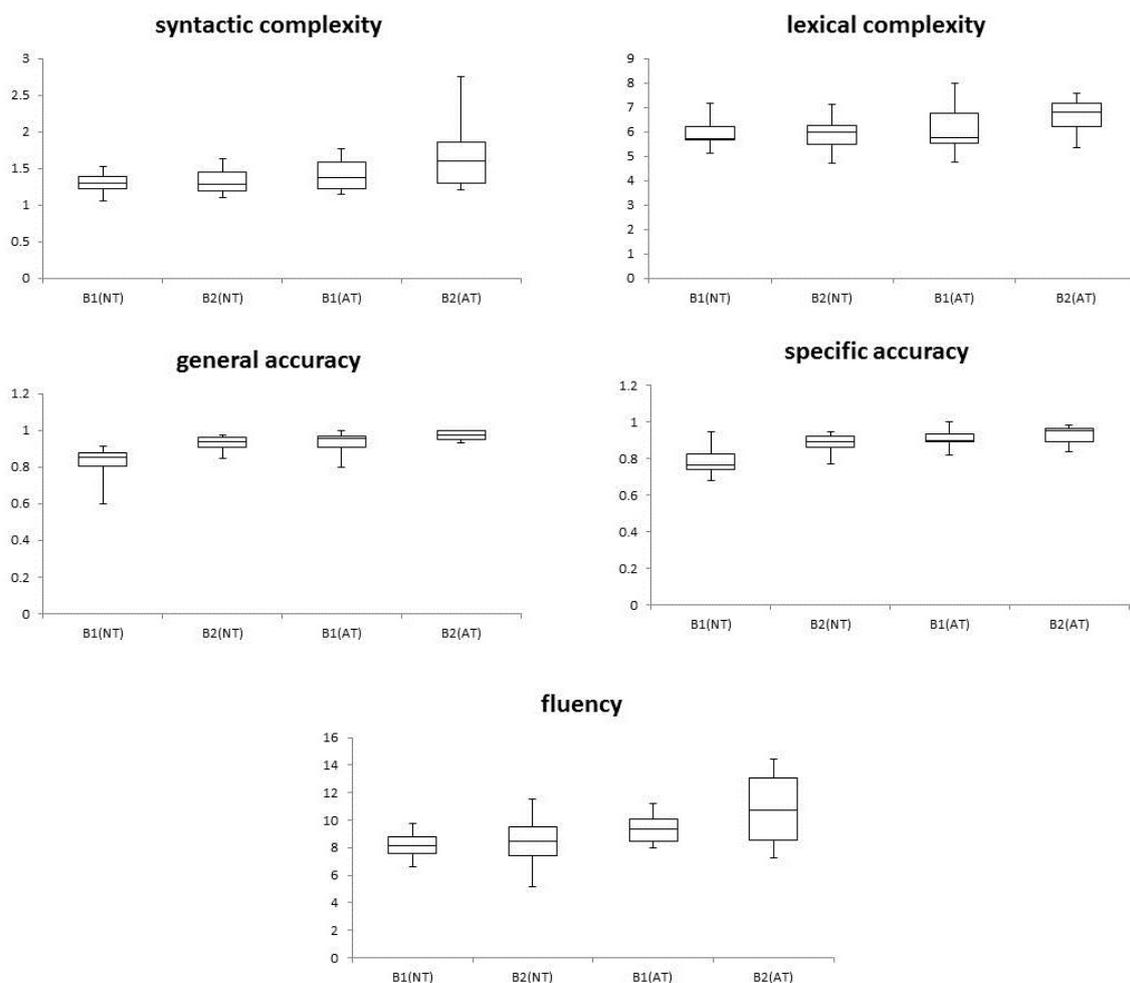


Fig. 2: Boxplots for each CAF variable results of both the narrative task (NT) and the argumentative task (AT) in CEFR level B1 and B2 groups: the CAF variables include syntactic complexity (the top left), lexical complexity (the top right), general accuracy (the center left), specific accuracy (the center right) and fluency (the bottom).

of results for each was obtained (Tables 5, 6), revealing that a significant task type effect (syntactic: $F(1, 28) = 17.30, p = .00, \eta^2 = .38$; lexical: $F(1, 28) = 9.50, p = .00, \eta^2 = .25$) was established without a significant L2 proficiency effect (syntactic: $F(1, 28) = 3.52, p = .07, \eta^2 = .11$; lexical: $F(1, 28) = 1.14, p = .29, \eta^2 = .04$) or a significant interaction effect (syntactic: $F(1, 28) = 3.75, p = .06, \eta^2 = .12$; lexical: $F(1, 28) = 2.08, p = .16, \eta^2 = .07$). More specifically, with regard to syntactic complexity, the B2 group ($M = 1.67$) performed better than the B1 group ($M = 1.42$) in doing the argumentative task. However, significant difference in the two proficiency groups was not found for the narrative task (B2: $M = 1.34$; B1: $M = 1.30$). Task-type had a significant effect on only the B2 group, as there was a significant difference in performance between the two tasks in the B2 group but not in the B1 group: B2 learners showed better

performance at the argumentative task than at the narrative task. On the contrary, in the case of lexical complexity, the argumentative task (B2: $M = 6.65$; B1: $M = 6.19$), which was irrelevant to L2 proficiency, triggered greater lexical complexity in the texts than the narrative task (B2: $M = 5.91$; B1: $M = 5.93$).

3.2 Accuracy

As shown in Table 7 and 8, the results of both general/specific measures for accuracy revealed significant effects of both L2 proficiency (general: $F(1, 28) = 15.92, p = .00, \eta^2 = .36$; specific: $F(1, 28) = 12.59, p = .00, \eta^2 = .31$), and task-type (general: $F(1, 28) = 62.22, p = .00, \eta^2 = .69$; specific: $F(1, 28) = 30.88, p = .00, \eta^2 = .52$). Also, the interaction between L2 proficiency and task type was significant (general: $F(1, 28) = 13.35, p = .00, \eta^2 = .32$; specific: $F(1, 28) = 6.32, p = .02, \eta^2 = .18$).

= .00, $\eta_p^2 = .18$). Therefore, a further simple main effects analysis was conducted, which indicated a statistically significant difference in general measures of accuracy between the argumentative task (B2: $M = 0.97$; B1: $M = 0.93$) and the narrative task (B2: $M = 0.93$; B1: $M = 0.82$) in both proficiency groups. In other words, both groups performed significantly better at the argumentative task than at the narrative task. Furthermore, both tasks enhanced the general accuracy of B2 level group more than the B1 level group. However, in terms of specific accuracy, the argumentative task (B2: $M = 0.93$; B1: $M = 0.91$) yielded greater verb-related accuracy in both groups, and for only the narrative task was the proficiency level of participants significantly correlated to their performance (B2: $M = 0.88$; B1: $M = 0.79$).

3.3 Fluency

As for fluency, the results showed that L2 proficiency did not have an effect ($F(1, 28) = 3.58, p = .07, \eta_p^2 = .11$) while task-type demonstrated a significant effect ($F(1, 28) = 43.53, p = .00, \eta_p^2 = .61$). Furthermore, a significant interaction was found between L2 proficiency and task-type ($F(1, 28) = 5.29, p = .00, \eta_p^2 = .16$). Post analysis revealed no statistically significant difference in fluency between both groups in the narrative task. However, the B2 group performed significantly better at the argumentative task ($M = 10.94$) than the B1 group ($M = 9.32$). In addition, the participants in both groups showed significantly greater fluency in the argumentative task, as compared to the narrative task (B2: $M = 8.56$; B1: $M = 8.17$).

4. Discussion and pedagogical implications

This study aimed at analyzing the effects of task-type on

Table 5: The results of two-way ANOVA regarding syntactic complexity

Source of Variance	SS	df	MS	F	η_p^2
L2 Proficiency (A)	0.31	1	0.31	3.52	0.11
Task Type (B)	0.78	1	0.78	17.30***	0.38
A*B	0.17	1	0.17	3.75	0.12

Note . *** $p < .001$

Table 6: The results of two-way ANOVA regarding lexical complexity

Source of Variance	SS	df	MS	F	η_p^2
L2 Proficiency (A)	0.75	1	0.75	1.14	0.04
Task Type (B)	3.8	1	3.8	9.50**	0.25
A*B	0.83	1	0.83	2.08	0.07

Note . ** $p < .01$

Table 7: The results of two-way ANOVA regarding general accuracy

Source of Variance	SS	df	MS	F	η_p^2
L2 Proficiency (A)	0.09	1	0.09	15.92***	0.36
Task Type (B)	0.09	1	0.09	62.22***	0.69
A*B	0.02	1	0.02	13.35**	0.32

Note . ** $p < .01$, *** $p < .001$

Table 8: The results of two-way ANOVA regarding specific accuracy

Source of Variance	SS	df	MS	F	η_p^2
L2 Proficiency (A)	0.05	1	0.05	12.59**	0.31
Task Type (B)	0.1	1	0.1	30.88***	0.52
A*B	0.02	1	0.02	6.32*	0.18

Note . * $p < .05$, ** $p < .01$, *** $p < .001$

Table 9: The results of two-way ANOVA regarding fluency

Source of Variance	SS	df	MS	F	η_p^2
L2 Proficiency (A)	15.07	1	15.07	3.58	0.11
Task Type (B)	47.1	1	47.1	43.53***	0.61
A*B	0.02	1	0.02	5.29*	0.16

Note . * $p < .05$, *** $p < .001$

L2 performance with Japanese EFL learners from two different proficiency levels by comparing performance on written argumentative and narrative tasks. Regarding the main research purpose (section 2.1), it could be concluded that both factors (i.e. task type and L2 proficiency) had a significant influence on L2 learners' writing performance. Here, we elaborate more specifically on the survey findings for each individual CAF dimension.

Syntactic complexity, as measured by the ratio of clauses per T-units, was significantly higher for the argumentative task than for the narrative task for only the B2 group. In the B1 group, no significant difference was found between the two tasks. As for the argumentative task the written texts in the B2 group were syntactically more complex than the ones in the B1 group; however, for the narrative tasks there was no significant difference in syntactic complexity between the two groups. With regard to lexical complexity (or lexical richness), as measured by the Guiraud index, is a measure which involves the number of different words used in a text. While the argumentative task significantly prompted both groups to use a more varied vocabulary than the narrative task, no significant difference in lexical complexity was found in the B1 and B2 groups for either task: it was not lexical complexity that discriminated these two proficiencies.

Two types of accuracy were analyzed, general and specific. In terms of general accuracy (i.e. involving global errors), both proficiency groups produced more accurate texts on the argumentative task than on the narrative task. In fact, B2 learners showed greater general accuracy than B1 learners regardless of the task. However, the two tasks did have different effects on the two proficiency groups with reference to specific or verb-related accuracy. While both groups performed better on the argumentative task than on the narrative task, L2 proficiency mattered for the narrative task: B2 learners used verbs more correctly in their narrative compositions but showed no difference on the argumentative task.

With respect to fluency (measured by the number of words per T-unit), differences were significant in favor of the argumentative task over the narrative task for both proficiency groups. There was a statistically significant difference in the fluency of the B1 and B2 learners on the argumentative task, albeit no significant difference was found on the narrative task. In other words, more proficient learners produced more fluent texts on the argumentative task.

In summary, the overall results showed a task-type effect on the L2 writing performance of Japanese EFL learners, regardless of their proficiency levels. More specifically, the argumentative task benefited writing

performance more than the narrative task. The higher L2 proficiency learners showed higher values on all the CAF dimensions when engaging in the argumentative task: i.e. they produced more syntactically complex, accurate, and fluent texts with a wider variety of vocabulary than they did on the narrative task. The results for the lower proficiency EFL learners exhibited a nearly identical pattern, with one exception: no significant difference in the two tasks was shown on the dimension of syntactic complexity. On the other hand, with regard to the L2 proficiency as a whole, the participants in the B2 group outperformed those in the B1 group on both tasks. On the argumentative task, significant differences between two proficiency groups on the dimensions of syntactic complexity, general accuracy, and fluency were established, while both measures of accuracy on the narrative task were significantly different for the two groups.

Task-types can differ from each other in terms of the degree of cognitive load which they place on EFL learners. In general, as compared to a narrative task, which typically offers some visual stimuli, an argumentative task is said to be more cognitively demanding⁶ in that EFL learners have to construct and develop their arguments logically based on critical thinking (assumed to involve a heavier cognitive load). This led to the prediction that the narrative task, often regarded as the less cognitively demanding task-type, would trigger higher values of all the CAF variables for both groups; however, in fact, our results indicated the reverse. Argumentative writing prompted a higher degree of complexity, accuracy and fluency in both groups. This may suggest that the element of familiarity is involved. Topic familiarity is also one of the major factors that can influence task complexity, exercising a great impact on L2 performance:²³ it can be expected that the higher familiarity yields greater complexity, accuracy and fluency. Given that topic familiarity varies in the amount of direct knowledge of topics, in this study, the argumentative writing had a subjective component in that it required the participants to write about themselves (i.e. *what do you do for your health?*): writing and opinion related to one's real-world personal experience could be more easy than describing an unfamiliar or unreal situation as given in the narrative task, thereby explaining the better performance on the argumentative essay. Thus, our findings could be explained by supposing that the participants were more familiar with the argumentative task than the narrative task, which eliminated some of the cognitive burden and allowed for "automatization" of the Japanese EFL learners' writing skill. In short, even if a task is

categorized as being of the same type as another, any task can embrace a wide degree of cognitive familiarity. Thus, taking into account the level of task familiarity ought to be a major goal of future research: it is of great significance to determine which factor can be attributed to a greater portion of Japanese EFL learners' writing performance. Moreover, care should be taken in drawing final conclusions based on limitations in our present study for the following reasons: (1) the 30 Japanese EFL learners were all intermediate level students; and (2) the sample-size was small with a limited range in proficiency levels. Further study should include a larger sample-size and cover a wider range of English proficiency levels. Furthermore, the use of CAF dimensions has been highly controversial in terms of reliability and validity. Although the CAF dimensions used in the study have proved to be some of the most reliable/valid domains for assessing L2 writing performance,⁷ measures used to assess CAF have varied in previous studies.

For these reasons, this study ought to be considered preliminary in nature and in need of further methodological refinement. Nonetheless, some pedagogical implications can be drawn from these findings. As mentioned above, different types of communicative tasks are primarily utilized in TBLT, encouraging language production, and task-based activities can help learners to reflect on both their own language and the target-language, resulting in a higher degree of cognitive awareness. It is the instructor's responsibility to provide effective tasks for students in EFL settings. In this study, the participants' writing performance varied depending on how much more or less cognitively demanding the task was; the less cognitively demanding the task was, the more their performance improved. Therefore it is highly recommended that students with lower proficiency should be first given less cognitively demanding tasks in order to enhance their fluency. More importantly, some explicit feedback with a focus on target-language forms is an essential requisite for making the best use of the student's cognitive awareness. This study revealed that for both tasks it was general accuracy that distinguished the two proficiency groups. The lower proficiency B1 level learners generally had more difficulty producing more accurate texts within the limited time available to complete the tasks, implying that they seemed to pay more attention to meaning than form. On the contrary, more proficient learners were able to pay attention to form and meaning, producing better results on the writing tasks. Thus, it can be expected that more direct benefit to the student could be provided through proficiency-level-specific feedback: in particular, detailed

grammatical feedback should be explicitly given to less proficient learners.

Competing Interests

Authors have declared that no competing interests exist.

References

1. Bygate, M., Skehan, P. & Swain, M. (2001). (Eds.). *Researching pedagogic tasks: Second language learning, teaching, and testing*, London: Longman.
2. Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.
3. Dickinson, P. (2010). Implementing task-based language teaching in a Japanese EFL Context. *University of Birmingham MA papers*. Retrieved on August 1, 2015, from <http://www.birmingham.ac.uk/documents/college-artslaw/cels/essays/languageteaching/dickinsonltdf>
4. Nakanishi, C. (2006). *A teaching approach to Japanese college students' EFL writing*. Tokyo: Keio University Press.
5. Dickenson, P. (2014). The Effect of Topic-Selection Control on EFL Writing Fluency. *Journal of Niigata University of International and Information Studies*, 17, 15-25.
6. 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.
7. Housen, A., Kuiken, F., & Vedder, I. (Eds) (2012). *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins Publishing Company.
8. Oxford University Press. (2001). *Quick placement test*. Oxford: Oxford University Press.
9. Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
10. Heaton, J. B. (1975). *Beginning composition through pictures*. Essex: Longman.
11. Siemens RG (1996). Lemmatization and Parsing with TACT Preprocessing Programs. *Computing in the Humanities Working Papers*. Retrieved on August 1, 2015, from

http://www.digitalstudies.org/ojs/index.php/digital_studies/article/view/78/111

12. JACET (The Japan Association of College English Teachers). (2003). *The JACET Lists of 8000 Basic Words*. Tokyo: JACET.
13. Hunt, K. (1996). Recent measures in syntactic development. *Elementary English*, 43, 139–732.
14. Storch, N. (2005). Collaborative writing: Product, process, and students' reflections. *Journal of Second Language Writing*, 14(3), 153-173.
15. Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47, 101-143.
16. Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510-532.
17. Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24 (4), 492-518.
18. Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17, 65-83.
19. 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, 59–84.
20. Burt, M. K., & Kiparsky, C. (1974). Global and local mistakes. In J. H. Schumann & N. Stenson (Eds.), *New frontiers in second language learning*. Rowley, MA: Newbury House.
21. Izumi, E., Uchimoto, K. and Isahara, H. (2004). *A Speaking Corpus of 1200 Japanese Learners of English*. Tokyo: ALC Press Inc.
22. Wolfe-Quintero, K., Inagaki, S., & Kim, H-Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity*. Honolulu: Second Language Teaching & Curriculum Center, University of Hawai'i.
23. Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.