Journal of Teaching Language Skills (JTLS)
35(4), Winter 2017, ISSN: 2008-8191
pp. 103-132

Task Complexity Manipulation and Accuracy in Writing Performance

Laleh Fakhraee Faruji
PhD Student
Science and Research Branch, Islamic Azad University, Teharn, Iran
fakhraeelaleh@yahoo.com

Farid Ghaemi*
Assistant Professor
Islamic Azad University, Karaj Branch, Alborz, Iran
ghaemi@kiau.ac.ir

Abstract

This study aimed to investigate the impact of task sequencing, along +/- reasoning demands dimension, on writing task performance in terms of accuracy. The study was motivated by Robinson's Cognition Hypothesis (CH) as well as previous studies investigating the relationships between task complexity and second language production. The participants of the study were 90 intermediate students at the Islamic Azad University, Shahr-e-Qods Branch, chosen from three classes based on their performance on the Preliminary English Test (PET). The participants in the three classes were assigned to three groups: Experimental 1, Experimental 2, and a Control group. At first, the students in all groups took part in the writing pre-test. Next, the treatment sessions including 8 sessions of picture description task performance began, during which the first experimental group received a series of picture description tasks based on a randomized order of cognitive complexity. The second experimental group received the same tasks, but ordered from simple to complex, based on their required reasoning demands. The control group, however, received some writing activities from the course book. Finally, during the last session, the post- test was administered to all participants. The results of the data analysis showed a significant positive impact for sequencing tasks from simple to complex on accuracy in writing task performance.

Keywords: accuracy, cognition hypothesis (CH), reasoning demands, task complexity, triadic componential framework (TCF)

Received: 01/08/2016 Accepted: 15/01/2017

* Corresponding author

During the last two decades there has been a growing interest in the use of tasks as a tool for language learning. Due to their potential to approximate L2 learners to performance in real life conditions, pedagogic tasks have been greatly concerned by SLA researchers, curriculum developers, teacher trainers, and language teachers worldwide.

Many researchers have investigated the use of tasks in relation to different language skills and components (e.g. Ellis, 2000, 2003, 2008, 2009a; Long, 1989; Long & Crookes, 1992; Nunan, 1989, 1991, 2004; Shehadeh & Coombe, 2012; Skehan, 1996; Skehan & Foster, 1997; Van den Branden, 2006, among others). However, despite the general enthusiasm for task-based approaches to pedagogy, there has been no consensus among researchers on the criteria based on which to sequence tasks in the syllabus.

Robinson (2001a, 2005, 2007b, and 2010) proposed a cognitively motivated solution to task sequencing by developing the 'Cognition Hypothesis' (CH). According to the claims of the Robinson's CH, "task sequencing, should be done by designing tasks simple on all relevant parameters of task demands first, and then gradually increasing their cognitive complexity on subsequent versions" (Robinson, 2010, p. 242).

Over the past two decades, there have been some studies investigating the CH through making comparisons between students' performance in two versions of a task; simple and complex (e.g. Gilabert, 2007; Iwashita, Mc Namara, & Elder, 2001; Kim, 2009; Kuiken & Vedder, 2007; Michel, Kuiken, & Vedder, 2007; Révész, 2009; Révész & Han , 2006; Soleimani & Rezazadeh, 2013; Steenkamp & Visser, 2011). Nevertheless, there has been paucity of research directed specifically at the effects of using a cycle of simple to complex tasks taking place over a longer period of time than a single classroom session (e.g. Robinson, 2007a; Thompson, 2014). Furthermore, most of the studies in this area have been done within oral modality, and relatively few studies have examined task complexity in relation to writing task performance (Kuiken & Vedder, 2007; Rahimpour & Hosseini, 2010; Salimi, Dadashpour, & Asadollahfam, 2011); therefore, this study aims at

developing this area of research by investigating the potential effects of sequencing tasks based on their cognitive complexity on L2 learners' writing task performance in terms of accuracy.

Literature Review

Robinson's Cognition Hypothesis (CH) and its related TCF

Robinson's Cognition Hypothesis (CH) (2001a, 2001b, 2003) made a distinction between two dimensions of task complexity; *resource-directing dimensions* and *resource-dispersing* dimensions. Resource-directing dimensions, direct learner's attention to particular linguistic features of a task, and resource-dispersing dimensions, deplete learner's attention over the different elements of the tasks.

An example of a *resource-directing* dimension is reasoning demands, where tasks do not demand reasoning from learners and just represent a simple transmission of information, require less conceptual and linguistic effort and less resources than a task with some reasoning demands. As far as *resource-directing* variables are concerned, Robinson (2001b, 2003, 2005, 2007b) argued that task complexity negatively affects fluency, but promotes accuracy and complexity. An example of a *resource-dispersing* variable is access to planning time during task performance; in that, giving no planning time increases the complexity of a task by simply dispersing attentional resources over the different aspects of the task. However, this dimension is also seen as important for syllabus design, as it prepares learners for real-life conditions, so practice along them "facilitates real-time access to an already established and also to a developing repertoire of language" (Robinson, 2003, p. 59). He predicted that increasing task complexity along resource-dispersing dimensions would have a negative effect on all aspects of L2 production.

Associated with the CH, Robinson developed the Triadic Componential Framework (TCF) (Robinson, 2001a). He distinguished three groups of factors, which interact to influence task performance including; "Task Complexity", "Task Difficulty", and "Task Condition";

of which Robinson (2001a, 2005, 2007b, and 2010) suggested complexity factors as the major basis for pedagogic task sequencing.

Task complexity factors are represented as "dimensions, plus or minus a feature, but can also be thought of in some cases as continuums, along which relatively more of a feature is present or absent" (Robinson, 2001a, p. 293). These dimensions of complexity, according to him, are "design features of tasks", and they can be manipulated to increase or lessen the cognitive demands of task performance. For example, tasks which require simple description of events happening now, in a shared context (+here and now), where few elements (+few elements) have to be described and distinguished consume less amounts of attentional, memory and reasoning resources than tasks which require reference to events that happened elsewhere (–here and now), in the past, where many elements have to be distinguished (–few elements), and where reasons have to be given to support statements made (+reasoning).

Reasoning demands as a variable of task complexity

As mentioned previously, Robinson (2001a, 2001b, 2005, 2007a, 2011) categorized reasoning demands as a part of the resource-directing dimension of task complexity. Getting incites from first language acquisition studies and psychological research, Robinson (2011) identified three aspects of reasoning, including: “spatial, intentional, and causal reasoning” (p. 15). In some studies, researchers attempted to distinguish different types of reasoning demands (spatial reasoning, i.e., reasoning about distance and position in physical space; intentional reasoning, i.e., reasoning about motives and intentions of people; and causal reasoning, i.e., reasoning about causes and effects of events); however, in other studies no distinction were made due to the fact that the distinction was very delicate, especially between intentional and causal reasoning.

There have been some studies which investigated the effects of task complexity, along various dimensions of cognitive complexity, on writing performance (e.g. Abdollahzadeh & Kashani, 2011; Frear, 2013;

Masrom, Alwi, & Daud, 2015; Rahimpour, 2007; Salimi, Alavinia, & Hosseini, 2012; Salimi & Dadashpour, 2012). However, of the small group of studies on cognitive task complexity and writing, the work of Choong (2014), Kuiken and Vedder (2007), and Frear (2013) manipulated task complexity along the same task complexity dimension as this study (i.e. reasoning demands). The following section briefly reviews these studies.

Kuiken and Vedder (2007) in their study conducted among 84 Dutch university students of Italian and 75 students of French, investigated the effect of task complexity on accuracy of L2 writing performance through manipulating task complexity along (+/-few elements and +/- reasoning demands). In their study, participants were presented with a prompt in L1 (Dutch) explaining that they had to write a letter to a friend regarding the choice of a holiday destination out of five options. In the letter a varying number of requirements had to be taken into account, six in the complex and three in the non-complex condition. In the complex condition a choice of a Bed and Breakfast in either Italy (for the students of Italian) or France (for the studentes of French) had to be made, while in the non-complex condition the writers had to choose a holiday resort in a country outside of Europe. The results showed that, using some global performance measures, a significant decrease of errors was observed in the complex task.

Frear (2013) investigated the effect of task complexity on two variables of the syntactic and lexical complexity of second language writing. . He manipulated task complexity along (+/- reasoning demands and +/- few elements) and examined its effect utilizing three separate letter-writing tasks (Task 1, low complexity; Task 2, medium complexity; Task 3, high complexity). Each task was designed to have different amounts of cognitive task complexity that were manipulated by increasing the complexity dimensions reasoning demands and number of elements in the task instructions. The results suggested that increases in cognitive task complexity might adversely affect dependant clause production, but benefit lexical production. However, Frear (2013) did not

investigate the effect of task complexity on accuracy of writing performance, which is the concern of this study.

Choong (2014) also examined the relationship between causal reasoning demands of tasks and the accuracy of written production of L2 learners. He used a video-retelling task which was a video-clip from a television show, in which, a character when preparing for a date, notices a wrinkle in his pants and it leads to a chain of events. The task comprised four prompts that differed in the amount of causal reasoning required to interpret the events told by the video, thus creating 4 different conditions. Causal reasoning was manipulated in terms of directness, intentionality, and agency in order to create conditions of no, low, moderate, and high causal reasoning demands. Through aggregate analyses, the study found that task complexity had an effect on the syntactic complexity of the participants' output, with task conditions 1 and 3 eliciting more complex output than task conditions 2 and 4. It was also found that measures of syntactic complexity have been unequally influenced by task complexity. When examining all the measures in aggregate by using Rasch analysis, the study found that task condition 2, the task condition which imposed the lowest level of causal reasoning demands (aside from task condition 1), elicited the least accurate output compared to the more complex task.

Different studies have manipulated different dimensions of task complexity as independent variables; however, most of the previous studies testing the CH had only addressed the immediate effects of manipulating task complexity variables; i.e. they involved one-off experiments that examined task complexity variables at a specific point in time (Ellis, 2005, 2009b). Consequently, there appeared to be few studies that had investigated manipulation of task complexity variables over time (Ellis, 2009b). This study is an attempt to fill this gap by manipulating task complexity using progressively cognitively complex versions of the same task type. The operationalization and manipulation of task complexity can also be transferred to pedagogic practice, syllabus design, and material development. In other words, teachers can receive

information regarding how learners are going to behave when facing the same task type differing in cognitive complexity.

Research Question

On the basis of the above literature review, the present study aimed at investigating the following research question:

Does manipulating task complexity along the resource-directing dimension (+/-reasoning demands) have any significant impact on EFL learners' writing task performance in terms of accuracy?

Method

Participants

Ninety homogenous students from three classes (thirty in each class) at Islamic Azad University, Shahr-e-Qods Branch, participated in the study. Their ages were between 20-32 years and the participants were both male and female. They were selected based on their performance on Preliminary English Test (PET). They were BA students majoring in English language translation. During the study, they were taking a two-credit Essay Writing course. They had studied English for 7 years in junior and senior high school levels before entering university. As a partial requirement of the university curriculum, all the Translation students had to participate in English Grammar I, English Grammar II, and Advanced Writing courses before taking the essay writing course. During the course, all the students studied the same text book "The practical writer," by Bailey and Powell (2008).

Instruments

To obtain the required data for this study, the following instruments were employed:

Preliminary English test (PET). The first instrument used in this study was an English general language proficiency test adopted from a Cambridge Preliminary English Test (PET, 2015); was an English

general language proficiency test to measure general language proficiency level of the participants and to ensure the homogeneity of the participants. The test available at http://www.cambridgeenglish.org/exams/preliminary/exam-format is mostly used for intermediate-level learners and is compatible with language proficiency level of most Iranian undergraduate learners majoring in English (Rahimi, 2009). It consists of three sections: reading and writing, listening, and speaking. The reading and writing section tests whether the participant can read and understand the main points from signs, newspapers and magazines, and can use vocabulary and structure correctly. The second section (listening) tests whether the participant is able to follow and understand a range of spoken materials including announcements and discussions about everyday life and finally, the last section (speaking) tests the participant's spoken English when they take part in conversation by asking/answering questions and talking, for example, about ones' likes and dislikes.

Pre-test and post-test. According to Skehan and Foster (1999) a way of ensuring that the task is reasonably demanding is to select a picture story that requires interpretation on the part of the participants. Hence, to fulfill the purpose of this study, a cartoon picture description task adopted from Abdollahzadeh and Kashani (2011) was used as both the pre-test and the post-test. According to them by being monologic rather than dialogic, the narrative-writing tasks can serve as a basis for deriving measures of learner performance that are not influenced by interactional variables. The selected task required participants to write a story based on a set of nine cartoon pictures. The picture story, although clearly structured with a chronologically ordered series of events, required interpretation on the part of the learners because the character's motive for performing different actions was uncertain until the final picture.

Picture arrangement (PA) subtest of Wechsler adult intelligence scale, revised version (WAIS-R). The Wechsler Adult Intelligence Scale (WAIS) is an intelligence test designed to measure cognitive ability in adults and older adolescents. The original WAIS was published in Wechsler (1955, as cited in https://www.wechsleradultintelligencescale.com/). It is currently in its fourth edition (WAIS-IV) released in 2008. The WAIS-R, a revised form of the WAIS, was released in 1981 and consisted of six verbal and five performance subtests. For the purpose of this study the Picture Arrangement (PA) subtest of WAIS-R was used to operationalize intentional reasoning demands. It consists of sets of pictures progressively increasing in the demands they make on the ability to reason about characters motives for, and intentions in, performing actions.

Procedures

At the beginning of the study, 113 students from three classes which were taking the two-credit Essay Writing course participated in the study. In fact, the classes had already been formed based on the University registration system. To ensure homogeneity of the participants, the researchers administered the PET, then the participants with extreme scores were omitted and 30 students in each class were chosen to participate in the study. To ensure the homogeneity of the subjects a test of statistical significance of the differences was used and it was shown that there was no significant difference between the participants in terms of their language proficiency (The details of the test are provided in the Results section). Then the homogenous students in three classes were assigned to three groups: Experimental A, Experimental B, and a Control group. In fact, other students were also receiving the treatments; however, their scores in the pretest and posttest were not considered in the data analysis.

During the first session, the students in all groups took part in the pre-test. A cartoon picture description task adopted from Abdollahzadeh

and Kashani (2011) was used as the pre-test. The participants were required to write a narrative account for the cartoon picture in thirty minutes. The time limit was set to ensure that the participants would use their time more efficiently.

From the second session, the treatment sessions began. However, it should be mentioned that the treatments were administered during the second forty five minutes of every class session. In fact, during the first forty five minutes of every session students in all groups received the writing lesson from the book entitled *The practical writer* by Bailey and Powell (2008), based on a pre-specified syllabus, and wrote a paragraph or an essay about an agreed upon topic.

During the second half of the class time, the first experimental group received a series of 8 picture description tasks in a randomized order of cognitive complexity; each task in one session. The second experimental group received the same tasks, but this time the tasks were sequenced from simple to complex based on their required reasoning demands; in fact, in the second experimental group the simplest task was administered in the first session and the most complex one was administered in the last session of the treatment. In other words, the students in the both experimental groups received the same tasks; however, the difference was in the order of tasks they received.

The control group, on the other hand, did not receive any picture description tasks; rather they received some writing activities from the course book which included some guided writing exercises related to what was taught in the first half of the class time. For example, during the first sessions, in which writing a well-organized paragraph was taught, the students were given a topic sentence and asked to find different types of supporting sentences and a reworded topic sentence to complete a paragraph, or they were given a paragraph with poor coherence and were asked to revise it for effective coherence. In later sessions, they were asked to read the sample essays in the book and outline their different parts.

During the second half of every class session within the experimental groups, first some task-related linguistic input, in the form of phrases that were identified to be helpful to complete the tasks, were provided to the students along with a set of related pictures for each task. In fact, in performing each task, the students in the two experimental groups were asked to view a set of pictures showing characters performing different actions, and decide which chronological sequence they should be arranged into in order to depict a coherent story. Then they were asked to provide a written description of the story that the series of pictures described (i.e., in the chronological order they had chosen). The students were given 30 minutes to arrange the pictures and provide a written description of the story during which, they were allowed to use a Persian to English dictionary.

Then the right arrangement of pictures and the correct description of the story was provided; therefore, the students became aware of their errors, and asked and answered questions regarding grammatical points and word choice. The questions were answered either by other students or the instructor.

Finally, during the last session, the post-test which was the same as the pre-test was administered to the participants in the three groups.

The treatment tasks. Getting insights from Robinson (2000, as cited in Robinson, 2005) reasoning demands was operationalized by using a series of one-way, closed picture arrangement tasks. Reasoning demand was differentiated by using a set of least to most complex picture sequences from the PA subtest of the WAIS-R. PA consists of ten tasks; the last eight of which were administered during the eight sessions of treatment.

In the PA subtest, sets of pictures progressively increased in the demands they made on the ability to reason about characters' motives for, and intentions in, performing actions. The simplest sequence consisted of three pictures depicting three stages, or successive actions, in the construction of a house, with no reasoning about the motives, intentions

or other thoughts of people. However, in the most complex version, pictures could only be successfully sequenced if motives, intentions and thoughts could be inferred. Based on Robinson's description of different types of reasoning; the tasks were designed to measure intentional reasoning.

Accuracy measurement in students' writing performance. Accuracy has been defined by Foster and Skehan (1996) as "freedom from error, based on whatever language is used" (p.304). Accuracy can be measured as a function of errors produced, either specifically (e.g., accuracy of verb forms) or generally (e.g., overall number of errors or error-free units) (Vercellotti, 2012). According to him general measurements of accuracy are better to be used when trying to measure general accuracy; on the contrary, specific measures of accuracy are better to be employed for research on a targeted structure but are less suited to capture overall accuracy performance. One of the general approaches to measure accuracy, according to Wolfe-Quintero, Inagaki, and Kim (1998), is to examine if a structural unit, such as a sentence, clause, or t-unit, is error-free. Due to the fact that this study did not focus on the accuracy of a specific targeted structure, one of the most common measures of general accuracy i.e. the number of error-free t-units per total number of t-units was used to fulfill the purpose of the study. In fact, following Wolfe-Quintero et.al (1998), the participants' general accuracy in writing in the pre-test and post-test was operationalized as the number of error-free T-units per T-unit. A t-unit was defined as a main clause plus any subordinate clauses (Hunt, as cited in Wolfe-Quintero et.al, 1998).

To ensure the reliability of the ratings, the students' accuracy in writing performance in the pre-test and the post-test in the three groups were rated by two raters; both the researcher and another expert teacher, and the researcher repeated the rating. Intra-rater reliability and inter-rater reliability with the second rater were calculated. The correlation between the two measurements done by the first rater ($r_{pretest}$= .993 and

$r_{posttest}$= .997) was significantly high. Therefore the two ratings done by the first rater enjoyed acceptable degrees of intra-rater reliability. Having established intra-rater reliability of the first rater's two ratings, the average of the two ratings was calculated and used for measuring inter-rater reliability of the ratings. The estimated nonparametric coefficient of correlation (Spearman's rho) between the ratings done by the two raters were also statistically significant ($r_{pretest}$= .978 and $r_{posttest}$= .990) hence, it was established that the two raters' ratings enjoyed considerable degrees of inter-rater reliability. During the whole study whenever writing scores were used in the analyses, the average of the three ratings done by the two raters was employed.

Results

The research question of the present study probed into the possible effect of independent variable on EFL learners' writing performance in terms of accuracy. In order to answer the research question, first the quality of the numerical data of the pre-test and post-test was analyzed with one-sample Kolmogorov-Smirnov tests of normality. Later, reliability of the estimates of accuracy was investigated and homogeneity of the samples in terms of their members' language proficiency and writing ability was evaluated.

Table 1 summarizes the descriptive statistics of the three groups in PET. The three mean scores pertaining to experimental group 1, experimental group 2 and the control group did not seem very different**.** Tables 2 shows the results of the test of statistical significance of the difference between the three groups in the PET.

Table1

The Descriptive Statistics of the PET of the Three Groups

	N	Minimum	Maximum	Mean	Std. Deviation
Preliminary English Test (PET) of the 1st experimental group	30	48.00	61.00	52.7	4.227
Preliminary English Test (PET) of the 2nd experimental group	30	42.00	60.00	51.9	3.977
Preliminary English Test (PET) of the control group	30	41.00	58.00	50.133	5.667

Table 2

Kruskal-Wallis Hypothesis Test of the PET Scores of the Three Groups

Hypothesis Test Summary

	Null hypothesis	Test	Test statistic	Sig.	Decision
1	The distribution of Preliminary English Test (PET) is the same across categories of group membership	Independent-Samples Kruskal-Wallis Test	2.053	.071	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

It was suggested by the results shown in Table 2 that the difference between the performances of the members of the three samples in the PET was not statistically significant.

Table 3 summarizes the descriptive statistics of the three groups in the pre-test. The three mean scores pertaining to experimental group 1, experimental group 2 and the control group (i.e. .51, .53, and .47) respectively.

Table 3

The Descriptive Statistics of the of the Pretest of Writing Accuracy of the Three Groups

	N	Minimum	Maximum	Mean	Std. Deviation
The raters' average rating of the pretest of writing for accuracy of the 1st experimental group	30	.16	.69	.5122	.12252
The raters' average rating of the pretest of writing for accuracy of the 2nd experimental group	30	.16	1.38	.5310	.25409
The raters' average rating of the pretest of writing for accuracy of the control group	30	.16	.70	.4734	.13440
Valid N (listwise)	30				

Comparing the writing accuracy of the participants of the three groups in the pretest was done with a Kruskal-Wallis test of significance of the differences. The logic for using this test was that the pretest scores were analyzed to be normal.

Tables 4 and 5 summarize the results of the Kruskal-Wallis test of the pretest of writing accuracy of the three groups.

Table 4

Kruskal-Wallis Test of the Pretest of Writing Accuracy of the Three Groups

Total N	90
Test Statistic	1.518
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.468
The test statistic is adjusted for ties.	

Table 5

Kruskal-Wallis Hypothesis Test of the Three Groups

Hypothesis Test Summary				
	Null hypothesis	**Test**	**Sig.**	**Decision**
1	The distribution of the pretest of writing accuracy is the same across categories of group membership.	Independent-Samples Kruskal-Wallis Test	.468	Retain the null hypothesis.
Asymptotic significances are displayed.		The significance level is .05.		

It was concluded from Tables 4 and 5 that there was no significant difference between writing accuracy of the pretest of the members of the two experimental groups and the control group.

Table 6 summarizes the descriptive statistics of the three groups in the post-test. The three mean scores pertaining to experimental group 1, experimental group 2 and the control group (i.e. .61, .76, and .53) respectively.

Table 6

The Descriptive Statistics of the of the Posttest of Writing Accuracy of the Three Group

	N	Minimum	Maximum	Mean	Std. Deviation
The raters' average rating of the posttest of writing for accuracy of the 1st experimental group	30	.33	.80	.6167	.11598
The raters' average rating of the posttest of writing for accuracy of the 2nd experimental group	30	.50	1.57	.7631	.19168
The raters' average rating of the posttest of writing for accuracy of the control group	30	.27	.77	.5313	.13329
Valid N (listwise)	30				

An ANOVA was used to compare the writing accuracy of the participants of the three groups in the posttest. The logic for using this test was that the posttest scores were proved to be non-normal.

The analysis of variance of the posttest of writing accuracy of the three groups pinpointed the difference between these sets of scores. Table 7, shows the Levene's test of homogeneity of variances of the posttest of writing accuracy of the three groups.

Table 7

Levene's Test of Homogeneity of Variances of the Posttest of Writing Accuracy of the Three Groups

Test of Homogeneity of Variances			
The Posttest of Writing Accuracy			
Levene Statistic	df1	df2	Sig.
.720	2	87	.490

Table 8 presents results of the ANOVA of the writing accuracy scores of the posttest of the members of the three samples in this study. The F (i.e. 18.194) was statistically significant as its level of significance (i.e. .000) was less than .05. This means that the members of the three groups had not equally improved their writing accuracy.

Table 8

Analysis of Variances of the Posttest of Writing Accuracy of the Three Groups

ANOVA					
The Posttest of Writing Accuracy					
	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	.824	2	.412	18.194	.000
Within Groups	1.971	87	.023		
Total	2.795	89			

However, the ANOVA did not specify where the difference between the samples lies. Therefore, the *post hoc* test employed in the multiple comparisons. Table 9 makes it apparent that while members of

the second experimental group in which picture description tasks were sequenced on the basis of their cognitive complexity outperformed members of the other two groups and were significantly better than them, there was no statistically meaningful difference between performance of the subjects from the other two groups in terms of writing accuracy.

Table 9

Multiple Comparisons of the Posttest of Writing Accuracy of the Three Groups

Multiple Comparisons

Dependent Variable: The Posttest of Writing for Accuracy

	(I) Group Membership	(J) Group Membership	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	Experimental Group 1 (Tasks used in a Randomized order of cognitive complexity)	Experimental Group 2 (Tasks sequenced based on their cognitive complexity)	-.14642*	.03886	.001	-.2432	-.0496
		Control Group (No picture description task)	.08533	.03886	.096	-.0115	.1821
	Experimental Group 2 (Tasks sequenced based on their cognitive complexity)	Experimental Group 1 (Tasks used in a Randomized order of cognitive complexity)	.14642*	.03886	.001	.0496	.2432
		Control Group (No picture description task)	.23175*	.03886	.000	.1350	.3285
	Control Group (No picture description task)	Experimental Group 1 (Tasks used in a Randomized order of cognitive complexity)	-.08533	.03886	.096	-.1821	.0115
		Experimental Group 2 (Tasks sequenced based on their cognitive complexity)	-.23175*	.03886	.000	-.3285	-.1350

*. The mean difference is significant at the 0.05 level.

In Table 10, the first experimental group and the control group were classified in the same subset because of the similarity observed between their members' writing accuracy in the posttest, and experimental group 2 was separated from the other two samples because of its members' outstanding writing accuracy in the posttest.

Table 10

Homogeneous Subsets of the One-way ANOVA of the Posttest of Writing Accuracy of the Three Groups

The Posttest of Writing for Accuracy				
			Subset for alpha = 0.05	
	Group Membership	N	1	2
Scheffe[a]	Control Group (No picture description task)	30	.5313	
	Experimental Group 1 (Tasks used in a Randomized order of cognitive complexity)	30	.6167	
	Experimental Group 2 (Tasks sequenced based on their cognitive complexity)	30		.7631
	Sig.		.096	1.000

Means for groups in homogeneous subsets are displayed.

a.Uses Harmonic Mean Sample Size = 30.000.

Based on the results presented in Tables 1 through 10, the null hypothesis was rejected. This is because the two experimental group members performed differently. In fact, outcomes of the study suggest that manipulating task complexity along reasoning demand positively effects EFL learners' writing accuracy provided that tasks are sequenced on the basis of cognitive complexity.

Discussion

The outcomes of the present study suggested that manipulating task complexity along with reasoning demand affected EFL learners' writing accuracy provided that tasks were sequenced from simple to complex on

the basis of cognitive complexity. While members of the second experimental group in which picture description tasks were sequenced on the basis of their cognitive complexity had outperformed members of the other two groups and were significantly better than them, there was no statistically meaningful difference between performances of the subjects from the other two groups in terms of writing accuracy.

According to Robinson (2003) manipulating task complexity along resource-directing dimensions (e.g. the amount of reasoning) may direct attentional and memory resources to task completion and therefore generate more accurate speech. Therefore, the findings of this study confirm the Robinson's cognition hypothesis within the written modality; in that, it provides further support for the positive impact of task complexity manipulation on EFL learners' accuracy.

Among different task complexity studies, Choong (2014), Frear (2013), Kuiken and Vedder (2007) and Sercu et al. (2006) manipulated task complexity along the same resource directing dimension as this study (i.e. reasoning demands). However, as it was mentioned earlier, none of these studies manipulated task complexity along a period of time; in fact, they manipulated task complexity in one-shot studies by providing the learners with two or more versions of the same task (with different degrees of complexity) during a single session. Except for this research there has been paucity of research directed specifically at the effect of a cycle of simple to complex versions of a task taking place over a longer period of time (e.g. Robinson, 2007a; Thompson, 2014); nevertheless, it is worth mentioning that both of these studies have been conducted in the oral mode of performance. In what follows, the results of this study will be compared with previous studies which have been closer to this study.

This study further supported the idea of following a Natural Order for sequencing that is the order of cognitive and linguistic development. It is possible to increase the cognitive demands of language learning tasks in parallel with the ontogenetic course of conceptual development in childhood, e.g. from tasks that require simple narrative description of

successive actions, with no causal reasoning to establish event relations, to those which require narrative description of simultaneous actions, and reasoning about participants' mental states (Lee & Rescorla, as cited in Robinson, 2003). Robinson stated that "such staged increases in the cognitive demands of tasks may therefore provide the learner with optimal, ontogenetically natural, contexts for making the form-function mappings necessary to L2 development" (p. 52).

In fact, the study showed that sequencing tasks based on their reasoning demands made the students engage multiple cognitive resources (attention and memory), which led to greater modification of problematic forms in their writing performance.

Confirming Robinson's (2001) CH, and being in line with Choong (2014) and Abdollahzadeh and Kashani (2011) studies, the results of this study clarified that sequencing tasks based on their cognitive complexity along the reasoning demands dimension could be a useful pedagogical practice for language teachers to make learners gradually develop their writing performance in terms of accuracy. However, the contrasting patterns of the findings with other studies (e.g. Kuiken and Vedder, 2007) might be related to the research design of these studies (i.e., different types of tasks used) and to the fact that different L2s and learners with different levels of proficiency were investigated.

Conclusions

This study was an attempt to investigate the impact of manipulation of task complexity along +/- reasoning demands on EFL learners' writing task performance in terms of accuracy.

The outcomes of the present study suggested that manipulating task complexity along reasoning demand affected EFL learners' writing accuracy provided that tasks were sequenced on the basis of cognitive complexity. While members of the second experimental group in which picture description tasks were sequenced on the basis of their cognitive complexity had outperformed members of the other two groups and were significantly better than them, there was no statistically meaningful

difference between performances of the subjects from the other two groups in terms of writing accuracy.

The results of this study may have some pedagogical implications:

The results of this study are of practical use for language teachers. Tasks can be sequenced gradually from simple to complex in order to develop learners' accuracy in writing. However, teachers should carefully consider the degrees of task complexity they are employing in their pedagogic tasks and its appropriateness for their learners' level of proficiency.

This study can also be applied in form-focused tasks. Different linguistic forms can be targeted and practiced using this procedure. However, with regards to matching tasks and language production there is not a direct, one-to-one relationship between language functions and linguistic forms. For example, there are many ways to express causality in English and many different methods may be employed by the participants; hence, it may be difficult to make predictions about language production according to task complexity. This may necessitate inquiry and examination by TBLT researchers regarding which task complexity dimensions, language forms and functions have strong relationships, and which may not (Choong, 2013).

Moreover, it can be argued that in order to facilitate inter-language development in L2 learners, L2 task designers need to consider the cognitive demands the tasks impose on the learners as a key point in their selection and sequencing of L2 tasks for both teaching and assessment purposes.

The findings of this study have also produced a number of possibilities for further research: First, the researcher in this study predicted about the causal processes that take place based on the linguistic performance data; however, in future research it will be better to gather independent evidence for the causal processes which might have occurred, instead of inferring based on linguistic performance data whether they have occurred; using introspective methods such as

stimulated recall procedure (Gass & Mackey 2000, as cited in Revesz, 2014).

Second, in conducting further research it will be better to further confirm that the task versions designed to be more complex than earlier ones, are indeed more cognitively demanding; by using different methods to obtain evidence for the required cognitive load to perform tasks. Some of these methods might be subjective self-ratings, subjective time estimations, secondary task methodology, and psycho-physiological techniques (Revesz, 2014).

Third, further research can be done in which task complexity manipulated along a specific resource-directing dimension, in both oral and written modalities simultaneously, which may allow for comparing the effect of one task complexity dimension in two modes.

Finally, further longitudinal task complexity studies need to be conducted to examine form-focused instruction to explore the development of specific linguistic forms which have been known to be difficult in L2 writing.

The goal of future research might be to address such limitations to pursue future research objectives, and thereby extend the potential significant theoretical, methodological, and pedagogical implications in L2 acquisition theory and practice.

References

Abdollahzadeh, S., & Kashani, F. A. (2011). The effect of task complexity on EFL learners' narrative writing task performance. *Journal of English Language Teaching and Learning*, *8*, 1-28.

Bailey, E. P., & Powell, P. A. (2008). *The practical writer with readings*. New York: Heinle & Heinle.

Choong, K. W. P. (2014). *Effects of task complexity on written production in L2English*. (Doctoral Dissertation, Teachers College, Columbia University). Retrieved February 2015 from http://gigalib.org/index.aspx

Ellis, R. (2000). Task-based research and language pedagogy. *Language Teaching Research*, *4*(3), 193-220.

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

Ellis, R. (2005). Measuring implicit and explicit knowledge of second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141-172.

Ellis, R. (2008). *The study of second language acquisition*. Oxford: OUP.

Ellis, R. (2009a). Task-based language teaching: Sorting out the misunderstandings. *International Journal of Applied Linguistics*, *19*(3), 221-246.

Ellis, R. (2009b). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics, 30*(4), 474–509.

Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition, 18*(3), 299-324.

Frear, M. W. (2013). *The effects of cognitive task complexity on writing*. Doctoral dissertation, Auckland University of Technology. Retrieved August 15, 2014 from http://aut.researchgateway.ac.nz/handle/10292/7309

Gilabert, R. (2007a). Effects of manipulating task complexity on self-repairs during L2 oral production. *International Review of Applied Linguistics in Language Teaching, 45*, 215-40.

Gilabert, R. (2007b). The simultaneous manipulation of task complexity planning time and [+-Here-and-Now]. In M.D.P. Garcia-Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 44-68). Clevedon, UK: Multilingual Matters.

Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information processing approach to task design. *Language Learning, 51*(3), 401–436.

Kim, Y. (2009). The effects of task complexity on learner-learner interaction. *System, 37*(2), 254–268.

Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics*, *45*(3), 261-284.

Long, M. H. (1989). Task, group and task-group interactions. *Unıversıty of Hawaii Working Papers in ESL, 8*(2), 1-26. Retrieved August 21, 2014 from http://files.eric.ed.gov/fulltext/ED366184.pdf

Long, M. H., & Crookes, G. (1992). Three approaches to task-based syllabus design. *Tesol Quarterly*, *26*(1), 27-56.

Masrom, U. K., Alwi, N., & Daud, N. M. (2015). The effects of task complexity on the complexity of the second language written production. *Journal of Second Language Teaching & Research*, *4*(1), 38-66.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics, 45*(3), 241–259.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555-578.

Nunan, D. (1989). *Designing tasks for the communicative classroom.* Cambridge: Cambridge University Press.

Nunan, D. (1991). Communicative tasks and the language curriculum. *TESOL Quarterly, 25*(2), 279-295.

Nunan, D. (2004). *Task-based language teaching*. Cambridge: CUP.

Preliminary English Test (PET) (2015). Retrieved March 3, 2016 from *http://www.cambridgeenglish.org/exams/preliminary/exam format*

Rahimi, M. (2009). The role of teachers' corrective feedback in improving Iranian EFL learners' writing accuracy over time: Is learners' mother tongue relevant? *Reading and Writing, 22*(2), 219-243.

Rahimpour, M. (2007). Task complexity and variation in L2 learners' oral discourse. *The University of Queensland Working Papers in Linguistics*, *1*, 1-9. Retrieved from https://www.library.uq.edu.au/

Rahimpour, M., & Hosseini, P. (2010). The impact of task complexity on L2 learners' written narratives. *English Language Teaching*, *3*(3), 198-205.

Révész, A. (2009). Task complexity, focus on form, and second language development. *Studies in Second Language Acquisition*, *31*(3), 437–470.

Révész, A. (2014). Towards a fuller assessment of cognitive models of task- based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics, 35*, 87–92.

Révész, A., & Han, Z. (2006). Task content familiarity, task type, and efficacy of recasts. *Language Awareness, 15*(3), 160 – 179.

Robinson, P. (2001a). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 287- 318). Cambridge: CUP.

Robinson, P. (2001b). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*, 27-57.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task- based language learning. *Second Language Studies, 21*(2), 45-105.

Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics, 43*(1), 1–32.

Robinson, P. (2007a). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, and perceptions of task difficulty. *International Review of Applied Linguistics in Language Teaching, 45*, 191-213.

Robinson, P. (2007b). Criteria for classifying and sequencing pedagogic tasks. In M. G. Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7-27). Clevedon: Multilingual Matters.

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In M. Putz & L. Sicola (Eds.), *Cognitive processing in second language acquisition: Inside the learner's mind* (pp. 243-268). Amsterdam: John Benjamins.

Robinson, P. (2011). Second language task complexity, the Cognition Hypothesis, language learning, and performance. In P. Robinson (Ed.), *Researching task complexity: Task demands, task-based language learning and performance* (pp. 3-38). Amsterdam: John Benjamins.

Robinson, P., Ting, S., & Urwin, J. J. (1995). Investigating second language task complexity. *RELC Journal, 26*(2), 62–79.

Romanko, R., & Nakatsugawa, M. (2010). Task sequencing based on the cognition hypothesis. In A. M. Stoke (Ed.), *JALT2009 Conference Proceedings* (pp. 436-445). Tokyo: JALT.

Salimi, A., Alavinia, P., & Hosseini, P. (2012). The effect of strategic planning time and task complexity on L2 written accuracy. *Theory and Practice in Language Studies*, *2*(11), 2398-2406.

Salimi, A., & Dadashpour, S. (2012). Task complexity and SL development: Does task complexity matter? *Procedia-Social and Behavioral Sciences*, *46*, 726-735.

Salimi, A., Dadashpour, S., & Asadollahfam, H. (2011). The effect of task complexity on EFL learners' written performance. *Social and Behavioral Sciences, 29*, 1390–1399.

Sercu, L., De Wachter, L., Peters, E., Kuiken, F., & Vedder, I. (2006). The effect of task complexity and task conditions on foreign language development and performance: Three empirical studies. *ITL-Review of Applied Linguistics, 152*, 55–84.

Shehadeh, A., & Coombe, C. A. (2012). *Task-based language teaching in foreign language contexts: Research and implementation*. Amsterdam: John Benjamins.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics*, *17*(1), 38-62.

Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185-211.

Soleimani, H., & Rezazadeh, M. (2013). The impact of increase in task cognitive complexity on Iranian EFL learners' accuracy and linguistic complexity: A test of Robinson's cognition hypothesis. *International Journal of Basic Sciences & Applied Research, 2*(5), 459-469.

Steenkamp, A., & Visser, M. (2011). Using cognitive complexity analysis for the grading and sequencing of isiXhosa tasks in the curriculum design of a communication course for education students. *Per Linguam, 27*(1), 11-29.

Thompson, C. (2014). *Guided planning, task complexity and second language oral development.* (Doctoral dissertation, University of Central Lancashire). Retrieved August 2014 from http://clok.uclan.ac.uk/10568/2/Thompson.

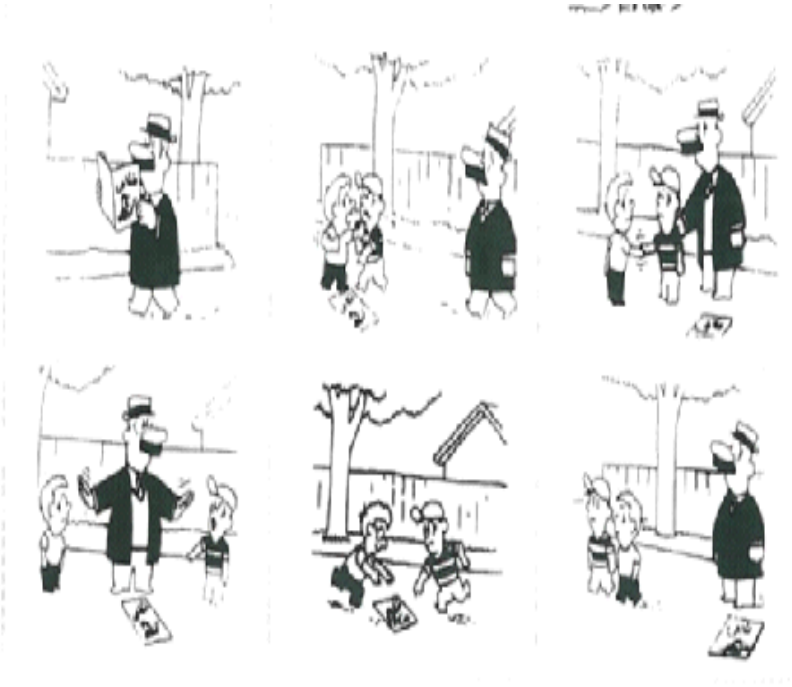
Van den Branden, K. (2006). *Task-based language teaching: from theory to practice.* Cambridge: CUP.

Vercellotti, M. L. (2012). *Complexity, accuracy, and fluency as properties of language performance: The development of multiple subsystems over time and in relation to each other* (Doctoral dissertation, University of Pittsburgh). Retrieved March 9, 2015 from www. http://d-cholarship.pitt.edu/12071/1/Vercellotti_CAF_v3.pdf

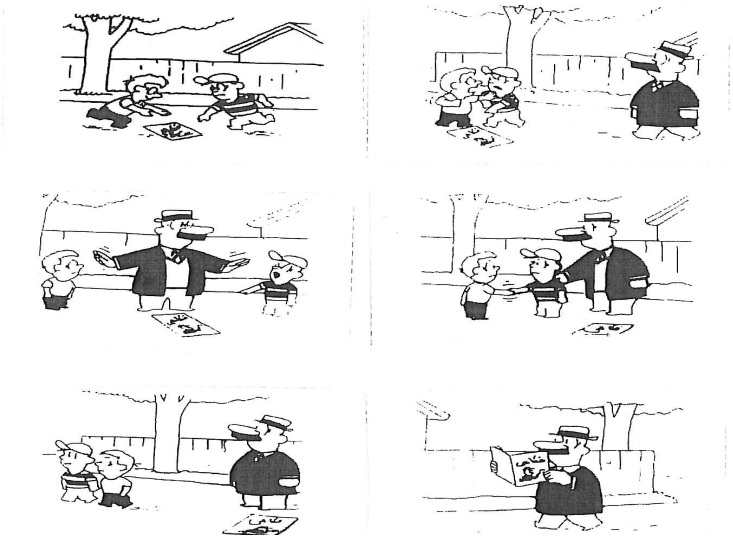
Wolfe-Quintero, K., Inagaki, S. & Kim H. -Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity.* Honolulu, HI: University of Hawaii Press.

Appendix: Two samples of the eight tasks

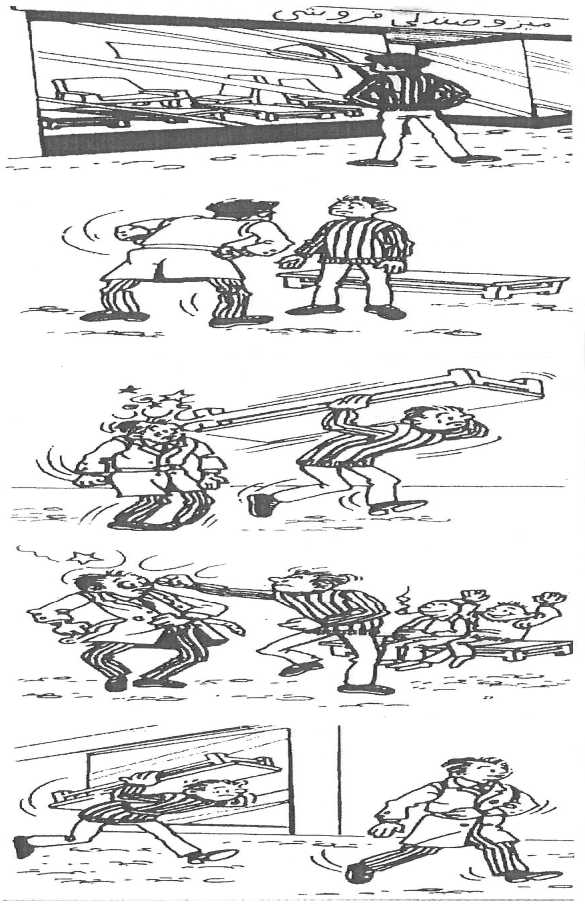
Task 4



Task 4 (correct order to make a coherent story)



Task 5



Task 5 (correct order to make a coherent story)