# Effects of Task Complexity on Written Performance of Low and Intermediate Proficiency Learners

BY

ABDULLAH ALSHEHRI

B.A., King Saud University, 2014

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Master of Arts in Linguistics
in the Graduate College of
the University of Illinois at Chicago, 2018

Chicago, Illinois

Defense Committee:

Susanne Rott, Chair
Xuehua Xiang
Richard Cameron, Advisor

I would like to dedicate this thesis to my parents,
and to my older brother, Faris.

# ACKNOWLEDGMENTS

In the name of Allah, the Merciful, the Compassionate, all thanks and praise are due to Him and peace and blessings upon His beloved servant and messenger, Muhammad.

I would like to first and foremost thank Dr. Susanne Rott. Your unrelenting patience has helped me through this process. I have sent you so many drafts, and with every draft, the rate of feedback increases. Your belief in my thesis has made it possible. I will always be grateful.

I would also like to thank Dr. Xuehua Xiang. Thank you for helping me walk the first steps into this journey. Know that this thesis began in your office. Thank you.

I thank Dr. Richard Cameron, whose class I needed to be in. I thank you for not accepting mediocre submissions and for not welcoming unsupported claims. Your teaching has guided me through the writing of this thesis.

My thanks and gratitude are also extended to Dr. Jessica Williams. Your guidance, advice, and patience were provided when most needed.

I thank Dr. Maja Grgurovic, from whom I learned many academic skills that were critically important in this thesis and in life. Only God knows how sincere you are in this job. You're so kind, it even escapes the notice of the kind. Thank you.

Section **Table of Contents** Page

# List of Tables

# LIST OF ABBREVIATIONS

| | |
|---|---|
| **CAF** | Complexity, Accuracy, and Fluency |
| **L2** | Second Language |
| **TBLT** | Task-Based Language Teaching |
| **CTTR** | Corrected Type-Token Ratio |
| **AoA** | Age of Acquisition |
| **EFCR** | Error-Free Clause Ratio |

## SUMMARY

Previous research on the effects of task complexity on performance has compared the claims of two models: Robinson's Cognitive Hypothesis (Baralt et. al, 2014; Robinson, 1996a; 2001; 2003a; 2011a; 2011b; Robinson and Gilabert, 2007a) and Skehan's Trade-Off Hypothesis (Skehan, 1996; 1998; 2014; Skehan & Foster, 1999; 2001). Robinson claims complex tasks will improve performance on measures of complexity and accuracy. Skehan, on the other hand, holds that if tasks are too complex a trade-off effect will occur between the constructs of complexity (syntactic & lexical), accuracy and fluency. Many studies examined the effects on oral production, but only a few treated written production; further, within the written production studies, only two studies examined task complexity effects on different proficiency levels in the same study. Therefore, the current study sought to determine the effects of task complexity on the written production of English as a Foreign Language (EFL) learners with an Arabic L1 background. Participants were two language groups: intermediate proficiency level group and low proficiency level group, A1 and B1, respectively, according to the Common European Framework of References for Languages (CEFR). The experiment divided each language group into two and gave one a simple task and the other a complex task. Data of a 127 passages were analyzed for syntactic & lexical complexity, accuracy, and fluency (CAF). Results of the low proficiency group indicate significant increases in syntactic & lexical complexity as well as fluency in the complex task, and no differences were found between the simple and complex tasks in accuracy. Results of the intermediate proficiency group indicate significant increases in all the CAF measures in the complex task. The difference in the results of both groups was attributed to study design limitations in the low proficiency group.

# I. INTRODUCTION

Task-Based Language Teaching (TBLT) has attracted many language teachers, syllabus designers and researchers. Teachers use tasks to foster use and acquisition of language, and syllabus designers use them to sequence curricula. In turn researchers study how tasks promote language use and uptake and try to come up with useful criteria for sequencing syllabi. Because tasks are "the potential building blocks of second language instruction" (Richards & Rodgers, 2001, p. 223), an increasing number of empirical studies have been conducted to test their design features and effects on language performance as well as on the acquisition of the language provided by performance, and on learning in general (e.g., Bygate et. al., 2008; Crookes & Gass, 1993; Ellis, 2003; Robinson and Gilabert, 2007a; Robinson, 1996a, 2011a, 2011b; Samuda and Bygate, 2008; Shehadeh & Coombe, 2012; Skehan, 1998; Thomas & Reinders, 2010; Van den Branden, 2006).

Numerous definitions of tasks were proposed by many second language experts (Long, 1985 p. 89; Nunan, 1989 p. 10; Prabhu, 1987, p. 24; Richards et. al., 1986 p. 289; Willis, 1990, p. 127). Ellis (2003), therefore, examined many of these definitions and came up with few criterial features of any task. He defines a task as including six features:

1. It is a workplan
2. It involves a primary focus on meaning
3. It involves real-world processes of language use
4. It can involve any of the four language skills
5. It engages cognitive processes
6. And it has a clearly defined communicative outcome

According to Ellis (2003) a task is a workplan designed from start to finish with all the steps in between, which could include the anticipation of specific linguistic structures,

vocabulary, or functions to be used by the learners. However, students may or may not use these anticipated lexical and grammatical aspects in a task. Instead of exclusively focusing on specific grammatical forms to use in achieving a task, a task seeks to engage students in meaning-making or in what Swain calls "languaging," which is the “process of making meaning and shaping knowledge and experience through language” (Swain, 2006, p. 98). Tasks could also emulate real-world processes of language use in the classroom, making them pedagogical tasks . Such tasks can involve reading, writing, listening or speaking. For example students can read a text and then summarize the main ideas to a classmate orally. Tasks also engage students in cognitive processes, such as "selecting, classifying, ordering, reasoning, and evaluating information" (Ellis, 2003, p. 10) while performing the task. Lastly, tasks have a non-linguistic goal which is used to determine the completion of the task. For example, in the example given, if the classmate understood the main ideas, then the task is achieved.

Tasks moved from being activities in already designed curricula and syllabi into being the basis on which curricula and syllabi are designed. Unlike views of linguistic structuralism, where syllabi are sequenced by linguistic structures, TBLT promotes tasks as the units of curriculum. The question in this line of research, then, becomes on what criteria should tasks be sequenced. Prabhu (1987) was the first person to suggest that complexity of tasks be the criterion. Inspired by Vygotsky's concept of 'zone of proximal development,' he argued tasks should pose a challenge that is attainable with some amount of effort. Prabhu suggested that teachers should assess complexity of tasks in relation to their learners' competencies and abilities in real-time. However, two limitations were pointed out by Baralt, Gilabert, Robinson, (2014) in this view, namely, that teachers' assessments would be first, impractical because teachers are required to continuously reassess tasks every time students perform them. Also such a measure is highly subjective because

teachers were not provided with a framework that operationalizes what complexity means exactly. Soon after, Long and Crookes suggested, in their 1992 seminal paper, that task sequencing decisions should be made in advance by conducting a needs analysis of target language use situations. As to the operationalization of complexity, they only mentioned what complexity might be based on briefly: "the number of steps involved, the number of solutions to a problem, the number of parties involved and the saliency of their distinguishing features..." (Long & Crookes, 1992, p. 44). However, Skehan (1996; 1998) was the first to actually introduce an organizational framework in this regard (Baralt et al., 2014). Although he agrees with Long and Crookes's views, his only precaution was that a great deal of emphasis on meaning-making and meaning conveyance might lead to sacrificing form. Coming from an information-processing perspective, he argued that language can be examined by three constructs: complexity, accuracy of form as well as meaning and fluency. He argued that his framework would make sure that a balance is achieved between the three constructs. Robinson, on the other hand, introduced the Triadic Componential Framework, which sought to establish in his words "a universal set of parameters" (Baralt et al., 2014) that can be used to measure the level of task complexity in order to inform task sequencing. Therefore, unlike Skehan, Robinson claims his framework is universal.

As for the effects of complexity on performance, two competing views have populated the literature: Robinson's Cognition Hypothesis and Skehan's Trade-Off Hypothesis. Robinson (Robinson, 2001a; Robinson & Gilabert, 2007) argues that both linguistic complexity and accuracy increases concurrently with task (cognitive) complexity (Baralt, et. al., 2014). On the other hand, Skehan (1998; 2014; Skehan & Foster, 2001) argues that as task complexity increases, a trade-off effect will occur between linguistic complexity and accuracy in particular, and also with fluency. Given that these two hypotheses have different views on how task complexity affects students'

performance, this study aims to test which view is more accurate. To do so, we engaged learners in a complex task as well as a simple task. We then measured their performance on four levels: lexical complexity, syntactic complexity, accuracy and fluency. In the following sections, we will first explain the Trade-Off Hypothesis, followed by the Cognition Hypothesis, and finally Robinson and Gialbert's Triadic Componential Framework (2007) will be elaborated on.

## II. TRADE-OFF HYPOTHESIS

Skehan's Trade-Off Hypothesis (1996; 1998) states that "humans have [a] limited information processing capacity and must therefore prioritize where they allocate their attention." He argues that because L2 learners' interlanguage is not yet proceduralized, both meaning and form need attention. As opposed to L2 learners, native speakers are able to perform complex tasks with greater ease and accuracy because their language is proceduralized, and so much of their attention is directed towards the task's content (meaning) (Skehan & Foster, 2001). Therefore, Skehan argues that if we provide learners with tasks that are too complex, much of their attention will be directed to its content, leaving little attention available to either be allocated to accuracy, linguistic complexity or fluency or a combination of two variables depending on the amount of attention left. Consequently, this little attention left will be directed to some of the aspects of performance (complexity, accuracy or fluency), and they will be developed. Other unattended to aspects, however, will be left underdeveloped. Therefore, if we follow a task-based syllabus with an ever-increasing complexity, students will start developing strategies to express the complex concepts without the chance to focus on grammatical forms. In turn, this will prevent interlanguage growth (Skehan, 1996, p. 41). Following this line of argument, Skehan emphasized the importance of designing tasks that create a balance between the three constructs of complexity, accuracy and fluency, so that all the aspects of performance are equally developed.

Skehan quotes a research study by VanPatten (1990) which sought to determine whether learners were able to pay simultaneous attention to form and meaning. VanPatten asked four groups of learners to listen to a short passage. The first group listened only to the content. The second group listened to the content and for one important lexical item. Groups three and four were asked to listen to the content as well as for non-communicative grammatico-morphlogical

forms (e.g., a definite article and a verb morpheme). His results showed that when students paid attention to language forms (Group 3 and 4), their comprehension of the passage was affected negatively. On the other hand, learners in group 2 were able to comprehend the passage while looking out for the single lexical item. VanPatten, therefore, concluded that students will have difficulty paying attention to both content and form. Skehan, hence, adopted the position that learners have a limited information-processing capacity, and will not be able to attend to both form and meaning in a cognitively complex task.

## III. COGNITION HYPOTHESIS

Robinson (1996a; 2001; 2003a; 2011a; 2011b; Baralt et. al, 2014; Robinson and Gilabert, 2007a), on the other hand, claims in his Cognition Hypothesis that by placing higher cognitive demands on students, students will be prompted to express these demands by using more complex linguistic forms, hence, linguistic complexity increases in more complex tasks. Furthermore, higher cognitive demands, to Robinson, means more attention directed to these linguistic forms produced, which in turn will render the production of these forms accurate. This entire process promotes interlanguage development in that if students succeed in providing the appropriate forms to achieve the cognitively complex task, form-function mappings will occur. On the other hand, if they fail to provide the appropriate forms, they will at least "notice" (Schmidt, 2010) that they are unable to express certain functions (e.g. comparing two items) or tenses (e.g. the past tense), which is the first step in the acquisition of these lacking aspects of the L2.

Claims of the Cognition Hypothesis are rooted in the view that human beings have multiple-resource pools of attention and that the capacity of each pool is unlimited (Robinson, 2003a). Robinson bases his view on two models of attention capacity and allocation. The first model is Delos Wickens's (2002) Multiple Resource Model, which holds that interference between two tasks occurs because of the shared resource pool they draw from. Hence, the more similar the used resource pools are, the more interference will occur. Furthermore, he posits that we have "multiple" resource pools, which means that we can draw from different resource pools without competition between them. Wickens proposes four categorical dimensions with two resource pools under each. Competition occurs only within each resource pool of each dimension. These dimensions are processing stages (perceptual/cognitive vs. response), visual

channels (ambient vs. focal), perceptual modalities (auditory vs. visual), and processing codes (spatial vs. verbal). As an illustration, this model explains that the reason why chatting with someone in the passenger seat while driving is easier than talking to them on the phone is because driving and using the phone both draw from the same resource pool: the spatial processing code (Wickens, 2002). Chatting with someone who is sitting in the passenger seat, however, is easier because it draws from another resource pool, the verbal processing code. The other model that influences Robinson's Cognitive Hypothesis is the Interference Model (Gopher, 1993; Sanders, 1998). The Interference Model claims that the competition occurring within a resource pool is not due to the limited capacity of the pool itself, but because there is not sufficient time allocated to the task. Therefore, it is the central executive of the working memory that is impacted here by a shortage of time, not that there is competition between pools nor that a particular pool capacity is limited. Robinson (2003a) combines the two models and argues that, first, individuals have multiple-resource attention pools, and that the attention pool has an unlimited capacity, hence, they are able to pay attention to both meaning and form.

## IV. TRIADIC COMPONENTIAL FRAMEWORK

Based on the Cognition Hypothesis, Robinson and Gilabert (2007) developed the Triadic Componential Framework (TABLE I) in an attempt to outline various task aspects that affect L2 learners' task performance. The framework is a taxonomy of pedagogic task design and classification that consists of three parameters: task complexity variables that are task-dependent; task conditions that are interactional (participation-related); and task difficulty variables that are learner-dependent.

**TABLE I**

**TRIADIC COMPONENTIAL FRAMEWORK**

| ***Task Complexity*** **(Cognitive factors)** | ***Task Condition*** **(Interactive factors)** | ***Task Difficulty*** **(Learner factors)** |
|---|---|---|
| (Classification criteria: cognitive demands) (Classification procedure: information-theoretic analyses) | (Classification criteria: interactional demands) (Classification procedure: behavior-descriptive analyses) | (Classification criteria: ability requirements) (Classification procedure: ability assessment analyses) |
| *(a) Resource-directing variables* making cognitive/conceptual demands | *(a) Participation variables* making interactional demands | *(a) Ability variables* and task-relevant resource differentials |
| +/− here and now<br>+/− few elements | +/− open solution<br>+/− one-way flow | [a] h/l working memory<br>[a] h/l reasoning<br>[a] h/l task-switching |

| | | |
|---|---|---|
| −/+ spatial reasoning<br>−/+ causal reasoning<br>−/+ intentional reasoning<br>−/+ perspective-taking | +/− convergent solution<br>+/− few participants<br>+/− few contributions needed<br>+/− negotiation not needed | [a] h/l aptitude<br>[a] h/l field independence<br>[a] h/l mind/intention-reading |
| *(b) Resource-dispersing variables* making performative/procedural demands | *(b) Participant variables* making interactant demands | *(b) Affective variables* and task-relevant state-trait differentials |
| +/− planning time<br>+/− single task<br>+/− task structure<br>+/− few steps<br>+/− independency of steps<br>+/− prior knowledge | +/− same proficiency<br>+/− same gender<br>+/− familiar<br>+/− shared content knowledge<br>+/− equal status and role<br>+/− shared cultural knowledge | [a] h/l openness to experience<br>[a] h/l control of emotion<br>[a] h/l task motivation<br>[a] h/l processing anxiety<br>[a] h/l willingness to communicate<br>[a] h/l self-efficacy |

[a] h/l= high/low

Starting from task conditions, these are factors that deal with the nature of participation in a task and the relationships between its participants, so whether negotiation of meaning is needed in a task or not, or whether the two participants share the same cultural background or not. This is not particularly important in our writing task because the audience to whom the learners are writing is unspecified. So, no interaction is happening between participants per se.

Second, Robinson differentiates between task complexity and task difficulty. Task difficulty refers to variables that learners bring to the task, such as the level of working memory

capacity in a learner, aptitude or the degree of willingness to communicate. These variables are unpredictable, subjective and learner-specific. Therefore, they are useful only in relation to a specific group of students. They are not very useful in sequencing tasks especially if we are trying to come up with a task sequence that is objective and can be applied in numerous contexts.

Lastly, task complexity is task-dependent, and so it is more objective. These variables can be used to raise the level of complexity of tasks regardless of learner individual differences. For example, if we ask students to talk or write about events that are happening in the present (+ here-and-now variable), it is more likely to be simple for them because the details of these events are still fresh in their memory, and so they are not recalling something from the past, nor are they visualizing events in the future; in parallel, the linguistic forms of expressing the here-and-now are simple. However, if we ask the same group of students to talk or write about events that happened in the past or will happen in the future, it will be more challenging because the details of what happened or will happen are not readily available in their memory, so recalling them from the past or creating them for the future are cognitively complex for all learners equally, so is expressing them in linguistic forms. For this reason, researchers and practitioners have focused on these "objective" task complexity variables to grade and sequence tasks.

Within task complexity, Robinson & Gilabert (2007) make an important theoretical distinction between what they call resource-directing variables that are cognitive and resource-dispersing variables that are performative. This distinction is what we alluded to earlier in that Robinson uses the term "task complexity" in a different light than Skehan does. Resource-directing variables are factors that make differential conceptual/linguistic demands on the learner. So, writing about events in the here-and-now or about events in the there-and-then place different cognitive demands (simple vs. complex, respectively). Another example would be the -

/+few element variable. In a task that asks learners to summarize a short story about a crime with four possible perpetrators (-few elements), students have to figure out who committed the crime first and give detailed justifications for each person in their summary; this is, therefore, a complex task. A simple task would be to read a short story with only one possible perpetrator (+few elements) and summarize it.

On the other hand, resource-dispersing variables are those that do not direct learner attention to certain concepts/linguistic forms, but put performative demands on the learner. Raising the performative demands means taking away scaffolding strategies and requiring native-like production. For example, a performatively simple task would provide students with planning time or with a task structure. A complex task would take away these variables, hence, making the task more real-world-like in its performance, not in its level of conceptual cognitive demands.

As for the interaction between the two, Robinson claims that raising the performative demands on a cognitively demanding task would decrease both linguistic complexity as well as accuracy. An example of that would be to ask students to write the summary of the crime story with the four possible perpetrators without providing time for planning.

This distinction between resource-directing and dispersing-variables is one not made by Skehan. He argues that both the variables that fall under cognitive complexity in Robinson's framework as well as the ones under performative complexity are both cognitive, and all cognitive variables to him will drain attention causing a trade-off in aspects of performance.

Therefore, the goal of this study is to test the different theoretical claims made by the two hypotheses regarding task complexity effects on performance in order to come up with the best way of sequencing task complexity. This will be conducted using Robinson's Triadic

Componential Framework to operationalize task complexity because unlike Skehan, he differentiates between two types of task complexity. Our indicators of performance are those of Skehan: complexity, accuracy, and fluency (CAF).

## V. LITERATURE REVIEW

Numerous studies have been conducted to examine which model is more accurate: the Cognition Hypothesis or the Trade-Off Hypothesis. The basic procedure followed by many studies is to divide a group of learners of the same proficiency level into two, and using Robinson's framework, assign a simple task to one of them and a more complex one to the other. Their performance is then measured in terms of linguistic complexity, accuracy and fluency. Previous studies have looked at both oral (Foster & Skehan, 1999; Ishikawa, 2008; Kim, 2009; Ortega, 1999; Rahimpour, 1997; 1999; 2007; Robinson, 1995; 2001; 2007; Skehan & Foster, 1999; Yuan & Ellis, 2003) as well as written task performance (Cho, 2015; Ishikawa, 2007; Kuiken & Vedder, 2008; Mohammadzadeh Mohammadabadi et al., 2013; Rahimpour, 2010; Ruiz-Funes, 2015; Salimi et al., 2011; Shajeri and Izadpanah, 2016)). However, since the current study looks only at the effect of task complexity on written performance, the review of literature will focus on studies that investigated written tasks.

Ishikawa (2007) explored the effects of task complexity on performance by manipulating the ±here-and-now variable. Learners in the simple task were asked to write narrative essays based on a picture sequence that required reference to events happening the present (+here-and-now). The complex task, on the other hand, required reference to events happening in the past, and was also based on a picture sequence; however, unlike the simple task, the picture was taken away after they consulted it (-here-and-now). He compared the accuracy, syntactic complexity, lexical complexity and fluency of the writings of 54 Japanese students. The results showed greater syntactic complexity, accuracy and fluency in the complex task. As for lexical complexity, the two types of measures Ishikawa used (lexical density and lexical variation), did not show statistically significant differences. Nevertheless, Ishikawa argues that because the

produced essays were lengthier (more fluent) in the complex task, maintaining the same level of lexical variation as the simple task would have meant that the production in the complex task was in fact more lexically complex. This is because as texts increase in length, their lexical variation automatically decreases. Therefore, maintaining the same level of lexical variation meant that it actually increased. Nevertheless, a limitation to Ishikawa's just presented argument is that the lexical variation measures he used were actually corrected for the increasing text effect. Earlier in the study, Ishikawa states that although the measures for lexical variation "'may not be perfect,' ...[they] do take into account the effect of text length." (Pilar, G. M., 2007, p. 143). So, in this regard, lexical complexity did not show significant differences between the simple and complex task in Ishikawa's study.

Similar to Ishikawa (2007), Hosseini & Rahimpour (2010) compared narrative essays manipulated by the ±here-and-now variable. The writings of 52 Iranian students with a Kurdish L1 were analyzed for accuracy, syntactic complexity and fluency. In their study design, the simple task encouraged the present tense and was context-supported while the complex task elicited the past tense and was context-unsupported. However, dissimilar to Ishikawa (2007), the only significant difference found in the study was the measure of fluency, where learners in the complex task produced longer T-units. The authors, therefore, concluded that their results supported Skehan's Trade-Off Hypothesis in that the complex task made students prioritize their attention and directed only to producing longer T-units (more fluent). Nevertheless, a trade-off effect did not occur in their study. Skehan argues that if aspects of performance increases, other aspects will decrease. As fluency increased in the complex task, linguistic complexity and accuracy maintained the same level as the simple task. If they had decreased, then we could have concluded that a trade-off effect occurred. Furthermore, the measure they used to measure

fluency, words per T-unit, is according to Lu (2017) and Vyatkina (2012), a measure of complexity. Vyatkina argues that "sentence length can be increased by two different types of complexification: adding more coordinate or subordinate clauses to a matrix clause or making clauses longer." In other words, the fact that T-units were longer in the complex task would either mean that learners used more subordinate clauses, which is an indicator of complexity, or that they had longer main clauses, which is also a measure of complexity. If this limitation is established, then the complex task, in fact, elicited more linguistic complexity, leaving fluency unknown.

Now, the revised results of Hosseini & Rahimpour (2010) study indicate that the complex task elicited more linguistic complexity while accuracy was maintained similar to the simple task. This conclusion, in fact, partially supports Robinson's Cognition Hypothesis. Robinson claims complex tasks increase both linguistic complexity as well as accuracy. Here, only linguistic complexity increased, while accuracy didn't support either hypotheses in that it neither increased nor decreased. This is because Robinson predicts it will increase, while Skehan's predicts it will decrease.

Similar inaccurate conclusions are observed in Salimi et al. (2011) study. They examined the written production of 29 Iranian female students with a Turkish L1 through the measures of accuracy, syntactic and lexical complexity, and fluency. The task they used were decision-making tasks manipulated by the ±few elements and ±reasoning demands variables. The complex task required of students to make decisions based on a story that had specific characters (e.g. pregnant woman) with connected and intricate relationships between them, and only one correct solution to be followed. The simple task, on the other hand, had similar characters with no relationships between them, and students had numerous solutions from which they could

choose. Their results showed significant increases in both complexity and fluency in the complex task with no differences in accuracy. Salimi et al. concluded that their results are in line with Skehan's Trade-Off Hypothesis. Nevertheless, similar with Hosseini & Rahimpour's (2010) study, they used the complexity measure, words per T-unit, to measure fluency. So, after correcting this conclusion, we end up with an increase in complexity with no differences in accuracy, which is the result that partially supports Robinson's Cognition Hypothesis.

Using the complexity measure of words per T-unit to measure fluency is best exemplified in Cho (2015). He studied the essays of a 110 Korean students in terms of accuracy, syntactic complexity and fluency. The tasks asked students to choose their preferred roommates from a list and write three reasons for their decisions. Manipulated along ±few elements and ±reasoning demands, the number of possible candidates as well as the number of properties for each candidate increased in the complex task. In the complex task, students had to choose four out of six candidates, each marked by six properties. In the simple task, students had to choose two out of four candidates, each marked with four properties. Results showed that task complexity manipulated along the ±few elements and ±reasoning demands variables did not show effects on complexity or accuracy, but students produced more fluent texts in the complex task. Therefore, the author concluded that his results supported Skehan's Trade-Off Hypothesis. Cho (2015) explains his conclusion stating that when students were given the (cognitively) complex task, they decided to allocate the limited attention they have to write more words, a measure of fluency. Nevertheless, in giving an example of how students' writing was more fluent, Cho states that "the complex group produced more words within a T-unit or sentence than [the] simple group by frequently making compound and complex sentences." This is enough evidence that the fluency measure is in fact a complexity measure. Also similar to the aforementioned studies, this

would mean that linguistic complexity was higher in the complex task with no significant differences found in accuracy; in turn, this means the results also partially support Robinson's Cognition Hypothesis.

Mohammadzadeh Mohammadabadi et al. (2013) studied the writings of 30 Iranian students of English through accuracy, syntactic complexity and fluency. He provided students with four narrative tasks based on a picture sequence. Two tasks were manipulated along the resource-directing factor of the ±here-and-now, and the other two were manipulated along the resource-dispersing factor of ±planning time. The ±here-and-now variable was operationalized in the use of the past tense in the complex task versus the use of the present tense in the simple task. The ±planning time was operationalized in the provision of a five-minute planning time in the simple task and with no planning time in the complex task. His results showed no significant differences in the tasks manipulated by the ±here-and-now variable. As for the tasks manipulated by the ±planning time variable, they showed higher accuracy in the planned (simple) task. Another limitation of this study is that, similar to Hosseini & Rahimpour (2010), the words per T-unit measure of complexity was used as a fluency measure. Nevertheless, because their conclusions in the resource-directing ±here-and-now variable did not show any significant differences, they did not support either hypothesis.

Looking at task complexity effects on two different proficiency levels, Ruiz-Funes (2015) analyzed essays of 32 learners of Spanish were through the measures of accuracy, syntactic complexity and fluency. Students were divided into 4 groups according to proficiency level and task complexity. Starting with low proficiency learners (based on the ACTFL proficiency guidelines), their complex task was an expository essay about the challenges of studying abroad, whereas the simple task asked for a personal narrative about the personal history and goals for

studying Spanish. In turn, high proficiency group engages in a complex writing task that asked for an argumentative essay discussing the problem of illegal immigration in Spanish. In contrast, the simple task engaged learners in an analytical essay discussing a Spanish-speaking country of choice. All in all, tasks were manipulated by ±familiarity of topic and genre, and ±reasoning demands. Results showed no significant differences, but an overall tendency towards higher syntactic complexity with decreases in accuracy and fluency in the complex task. Hence, the author claims the results supported Skehan's hypothesis that the complex tasks pushed participants in the study to attend to aspects of language performance at the expense of others. His results also showed that proficiency level manipulation was not significant between the intermediate- and high-proficiency students.

In measuring syntactic complexity and fluency, Ruiz-Funes (2015) used measures that were somehow more comprehensive than other researchers. For syntactic complexity, he used measures that looked at length of production, coordination, and subordination. For fluency, he used length of text as the measure because the tasks were timed to 50 minutes. For the current investigation, we adopted the aspects of syntactic complexity Ruiz-Funes (2015) used: length of production, coordination, and subordination, as well as the exact measure of fluency due to the similarity of our tasks conditions (timed tasks).

Finally, Kuiken and Vedder (2008) looked at the effects of task complexity on written performance in two different proficiency levels. They studied the written production of 91 Dutch students of Italian and 76 students of French through the measures of syntactic complexity, lexical variation and accuracy. Students were asked to write a letter to a friend suggesting a vacation destination. Task complexity was manipulated along the ±few elements variable, in which the complex task made students consider six options before choosing a destination, such

as the presence of a garden or a place to do physical exercise. In contrast, the simple task required students to consider three options only. Results showed a significant increase in accuracy in the complex task, while no differences were found in both syntactic complexity and lexical variation between the tasks. Furthermore, the results showed no significant differences in the results of both proficiency levels in that the same results were observed in both proficiency groups. The authors, therefore, concluded that their results are in partial support of Robinson's Cognition Hypothesis, and that there is not interaction between language proficiency and task complexity.

The studies reviewed above reveal several limitations. Starting with the ones related to the CAF measures, two in particular seem to arise. 1) First, a substantial number of previous studies (Cho, 2015; Hosseini & Rahimpour, 2010; Mohammadzadeh Mohammadabadi et al., 2013; Salimi et al., 2011) violated the validity of measures in that they used a measure for complexity to assess fluency; namely, they used the words per T-unit measure of complexity to measure fluency. 2) This in turn, led to reaching incorrect conclusions. Using incorrect measures caused a number of studies (Cho, 2015; Hosseini & Rahimpour, 2010; Salimi et al., 2011) to argue that their results supported Skehan's Trade-Off Hypothesis; however, when the measure was corrected, their results turned out to be in partial support with Robinson's Cognition Hypothesis. 3) The third limitation was that three studies used a single measure to capture each construct (complexity, accuracy and fluency) (Hosseini & Rahimpour, 2010; Mohammadzadeh Mohammadabadi et al., 2013; Salimi et al., 2011). As Norris and Ortega (2009) argue: the claim that the "CAF constructs must be measured multidimensionally is not only theoretical, but also empirical." They meant that each construct (complexity, accuracy and fluency) needs to be looked at from different dimensions. So, if we take syntactic complexity as an example, it can be

observed through the increased use of coordination, subordination, length of production, amongst many other dimensions. Furthermore, Kim, Crossley, and Kyle (2018) studied the different measures denoting lexical complexity. They concluded that lexical complexity is a "multidimensional phenomenon" that needs various measures to be accurately and comprehensively captured. Therefore, using more measures to capture the different aspects of each construct would more accurately and comprehensively delimit the constructs. 4) The fourth limitation was that many studies defined complexity as merely syntactic, overlooking lexical complexity (Cho, 2015; Hosseini & Rahimpour, 2010; Mohammadzadeh Mohammadabadi et al., 2013; Ruiz-Funes's, 2015). The literature, therefore, lacks a clear understanding of the effects of task complexity on lexical complexity.

Furthermore, the findings of the studies are conflicting. Some studies showed no significant differences in all the measures (Ruiz-Funes's, 2015; Mohammadzadeh Mohammadabadi et al., 2013), others showed an increase in linguistic complexity with no change in accuracy (Cho, 2015; Hosseini and Rahimpour, 2010; Salimi et al., 2011), one of the studies showed an increase in accuracy with no change in linguistic complexity (Kuiken & Vedder, 2008), lastly, one study showed (Ishikawa, 2007) significant increases in all measures except lexical complexity.

Henceforth, given the limitations pertaining to the CAF measures, we have taken great care in choosing measures that were proven valid in the literature in order to avoid reaching the aforementioned invalid findings. Second, aside from fluency, we use a number of measures for each construct, as will be mentioned later. Third, we used seven measures to capture lexical complexity because of the need for it in this line of research; however, this will not be at the expense of the other constructs. Lastly, the interaction of proficiency levels and task complexity

effects on written performance were looked at in only two studies (Kuiken & Vedder, 2008; Ruiz-Funes, 2015); both of them nullified the existence of an effect. Therefore, this study also aims to either confirm previous conclusions or bring about new results.

## VI. RESEARCH QUESTIONS

1. What are the effects of task complexity on the written production of low proficiency learners in terms of complexity, accuracy and fluency?

2. What are the effects of task complexity on the written production of intermediate proficiency learners in terms of complexity, accuracy and fluency?

## VII. METHODS

A. Participants:

Participants in the study belonged to two proficiency levels (low and intermediate). Learners were placed in the low level after taking an online placement test prepared by Oxford University Press and were considered level A1 on the Common European Framework of Reference for Languages (CEFR). The second group was an intermediate level group, and those were placed at the A1 level a year earlier; they had studied at the university for a year and were considered level B1 (CEFR) at the time of data collection. Each language group was divided into two; a simple task was assigned to one of the two subgroups and a complex task was assigned to the other. So, in total, we had four groups: low proficiency level group with a simple task and another with a complex task; intermediate proficiency level group with a simple task and another with a complex task. Seventy two participants were in the low proficiency level (57 in the simple task and 15 in the complex task), and fifty five in the intermediate proficiency level (29 in the simple task and 26 in the complex task.

### B. Tasks:

#### 1. Low Proficiency Learners Tasks

The tasks (APPENDIX A) assigned for the low proficiency learners were manipulated by the +/-causal reasoning variable (look at Framework above). The simple task was categorized as simple because it required no causal reasoning. Students were asked to describe their daily routine: what they eat, drink, when they do things, and how they enjoy leisure time. This type of task evidently requires simple information transmission, where students spell out their routine. On the other hand, the complex task was categorized as complex because it involved causal

reasoning. It asked students to write about an important person in their lives and explain why this person is important. The fact that it asked "why," made us decide that the writing assignment was more cognitively complex.

As for the rest of the resource-directing variables (see TABLE I), both tasks were equal in that they both asked about few clearly distinguishable elements in the here and now. These few elements did not require perspective-taking, in which an issue is treated from different perspectives nor did it require spatial reasoning, where task success requires the description of a location, nor intentional reasoning, where characters' intentions are talked about.

Lastly, both tasks needed to be controlled in regards to the resource-dispersing variables so as to not diminish task complexity in one task as opposed to the other (according to Robinson, TABLE I). Students in both tasks were given a single task (as opposed to dual-task) with a clear structure to be followed. The steps in the structure are independent of each other, in that skipping a step causes no harm to task achievement, for example if a learner didn't talk about what he drinks on a daily basis, although the structure asks him to write about that, it didn't harm task achievement. Lastly, students were not allotted time for planning, nor were the tasks familiar, and knowledge about the tasks was not provided prior to or during task performance, for example a task with provided information would give the learners names of different foods, activities, or time expressions to be used in the task.

### 2. Intermediate Proficiency Learners Tasks

The tasks (APPENDIX A) in the intermediate proficiency group were manipulating the +/-here-and-now variable (look at Framework above). The simple task was simple because it required language learners to write about a book or a TV program that they liked. The task was

descriptive and was cognitively simple because it mainly involved talking about something in the here-and-now. Linguistically this task was presumed to mostly elicit the use of the present tense. In turn, the complex task required learners to write about their future plans after graduation. Unlike the simple task, where learners had knowledge of what they were writing about, the complex task was cognitively demanding because it required learners to think about the far future and, in more technical terms, initiate future episodic thinking. The complex task was also categorized as cognitively complex for reasons outside of Robinson's framework. Learners whose tests we analyzed were in their first year of college, so the likelihood of thinking about what they are going to do after graduation, that is after another four years, is slim. This is evidenced in memory studies that examined the content of future-oriented thinking. They show that future-oriented thinking predominantly revolves around short-term concerns (D'Argembeau & Van der Linden, 2004; D'Argembeau et al., 2009; Klinger, 1971; Spreng & Levine, 2006).

As for the other resource-directing variables, both tasks contained few elements that did not require perspective-taking, nor spatial or intentional reasoning. The tasks were also equal in reference to the resource-dispersing variable. They both consisted of a single structured task with a few independent steps: three steps in each task (APPENDIX A). Planning time was not provided nor was background information about the tasks. TABLE II presents a summary of the conditions.

**TABLE II**

SUMMARY OF TASK COMPLEXITY AND PROFICIENCY LEVELS

| | **Proficiency level** | **Task complexity** | **Resource-directing variable** |
|---|---|---|---|
| **1** | Low | Simple | No causal reasoning |
| **2** | Low | Complex | Yes causal reasoning |
| **3** | Intermediate | Simple | Yes here-and-now |
| **4** | Intermediate | Complex | No here-and-now |

## VIII. PROCEDURE

The data were 127 timed (50 min) passages written by Saudi students who are native speakers of Arabic. The data was collected in an EFL setting as their final exams in a public university. Passages were rated by instructors who have a minimum of an MA in TESOL and at least 2-5 years of experience. Passages were hand-written, and so we first typed them into a Microsoft Word file before analysis. Each passage was analyzed for the following three constructs: linguistic complexity (syntactic and lexical), accuracy and fluency. TABLE III summarizes the measures we used in the analysis of these passages categorized according to the construct. In the next sections, we will explain the detailed procedures followed in analyzing each construct as well as the reasoning behind choosing each measure.

**TABLE III**

THE MEASURES USED TO ANALYZE COMPLEXITY, ACCURACY AND FLUENCY

| Constructs | | Measures | Explanation |
|---|---|---|---|
| ***Syntactic complexity*** | **1** | **T-units / sentence** | T-units per sentence (coordination) |
| | **2** | **Clauses / t-unit** | Clauses per T-units (subordination) |
| | **3** | **Clauses / Sentence** | Clauses per sentence |
| | **4** | **Words / Clause** | Words per clause |
| ***Lexical complexity*** | **1** | **CTTR** [a] | Corrected type-token ratio |
| | **2** | **Bigram Proportion** | Proportion of native-like Bigrams |
| | **3** | **Trigram Proportion** | Proportion of native-like Trigrams |
| | **4** | **Familiarity** | Words on a scale of Familiarity |
| | **5** | **Concreteness** | Words on a scale of Concreteness |
| | **6** | **AoA** | Words on a scale of Age of Acquisition |

| | | | |
|---|---|---|---|
| | **7** | **Word Specificity** | Words with specific meanings |
| ***Accuracy*** | **1** | **Errors / clause** | Errors per clause |
| | **2** | **Errors / T-unit** | Errors per T-unit |
| | **3** | **EFCR** [b] | Error-free clause ratio |
| | **4** | **Errors / Words** | Errors per words |
| ***Fluency*** | **1** | **W** | Total number of words |

[a] CTTR: ratio of different words (type) to total number of words (token) corrected for text length effect

[b] EFCR: ratio of error-free clauses to total number of clauses

### A. Syntactic Complexity Criteria

In order to measure syntactic complexity, we needed to identify sentences, T-units and clauses first. Since the essays were hand-written by learners who were low and intermediate in proficiency level, identifying sentences, T-units and clauses was especially challenging. Therefore, criteria were followed in order to make the analyses more reliable, and we explain these under the subheadings below.

#### 1. Sentence

Although punctuation and capitalization were not taken into consideration in measuring accuracy, they were deemed useful in recognizing the parameters of a sentence. When learners ended a sentence with a period and started another with a capitalized letter, they were considered two sentences; for example, "I get up at 6 in the mourning. I go to university at 7 in the

mourning." If a learner wrote a period at the end of a sentence and started another with a small letter, they still were considered two sentences; for example, "Then I go to university at 9:00 am. every day I go to bed at 10:00 pm..." Lastly, if a learner did not end a sentence with a period but started another with a capitalized letter, they still were considered two sentences; for example, "They are the important persons in my life They help me..." These three rules were followed as much as possible. If these rules did not apply, we used each learner's style of writing in the L2 and context to determine the parameters of sentences.

**2. Clause**

In defining a clause we followed Kellogg Hunt (1965) who defined a clause as "a visible subject and a finite verb" (p. 29). Clauses are divided into 1) independent clauses, which are clauses with a subject and a predicate and is a complete thought and can stand on its own as a sentence. 2) Dependent clauses are clauses that have a subject and a predicate but are not a complete thought, hence, cannot stand on its own. We counted independent clauses by the occurrence of a coordinator between two sentences and independent clauses by the occurrence of a subordinator before the subordinate clause.

**3. T-unit**

A t-unit is the smallest grammatically allowable unit, which could consist of a single clause or a main clause and its subordinate clause. Hunt (1965) defines it as a "minimally terminable unit." To count T-units, we counted subordinate clauses and subtracted them from the number of all clauses. To illustrate the validity of our method we will give an example. Say we have a text with ten sentences; five of these sentences have coordinate clauses next to the main

clauses, and the other five have five subordinate clauses next to the main clauses. So, in total we have 20 clauses (10 main clauses, 5 coordinate clauses and 5 subordinate clauses). If a T-unit is the smallest grammatically allowable unit, then a subordinate clause doesn't stand by itself; therefore, we subtracted the number of subordinate clauses (5) from overall clauses (20), resulting in 15 independent clauses. We applied this method to all of the passages.

**B. Syntactic Complexity Measures:**

In the current investigation syntactic complexity was measured by the length of clauses, the number of clauses, coordination, and subordination. The use of clause length as a measure of complexity was motivated by previous research that found length of production measures, in general, to be useful in showing syntactic complexity (Henry, 1996; Ishikawa, 1995; Larsen-Freeman, 1978; Ortega, 2003). Length of clauses, in particular, was found by Lu (2011) to be the best indicator of complexity. He studied fourteen complexity measures grouped into 5 types, and found that out of the length of production type of measures, length of clause was the best in that it showed complexity between adjacent proficiency levels. Furthermore, Vyatkina (2012) conducted a longitudinal study of language development and used length of clauses as one of two complexification methods, next to adding more coordinate or subordinate clauses. Her results showed that length of clause was an effective measure of complexity. The second measure we used was the clauses per T-unit measure, which measures the amount of subordination (Lu, 2017). This measure was used in numerous task complexity studies (Cho, 2015; Ishikawa, 2007; Kuiken &Vedder, 2008; Sasayama, 2011; Shajeri & Izadpanah, 2016) with significant results in some of them (Ishikawa, 2007; Sasayama, 2011). Next, the T-units per sentence measure looks at complexity in the amount of coordination (Lu, 2017). Similar to the

previous measures, recent studies also used this measure in examining task complexity effects on performance. (Cho, 2015; Ruiz-Funes, 2015) However, no significant results were found in the two referenced studies using this measure. Lastly, we used a measure that was categorized as general in Lu (2017), which is the clauses per sentence. We used this measure it did not look at a specific aspect of syntactic complexity but treated complexity as a whole.

### C. Lexical Complexity and Fluency Criteria

Unlike syntactic complexity and accuracy, lexical complexity and fluency were analyzed using natural language processing tools. First, we used the freely available Laufer and Nation's Lexical Frequency Profile (LFP) software to examine the Corrected Type-Token Ratio and the total number of words. The LFP replaces contractions by constituent words, eliminates punctuation and single letters except for 'a' and 'I.' The second tool we used was Crossley and Kyle's (2015) Tool for the Automatic Analysis of Lexical Sophistication (TAALES). We used this tool to analyze the rest of lexical complexity measures, which are the n-gram proportions (bi-gram & tri-gram), familiarity, concreteness, AoA, and word specificity. TAALES is a freely available language analysis tool that examines lexical sophistication, and it incorporates more than 484 measure. Before inputting texts into the two software, spelling errors needed to be corrected.

### D. Lexical Complexity Measures

Lexical complexity was measured using seven measures. Starting with the corrected type-token ratio (CTTR), it is a measure that calculates the ratio of different words in a text to the total number of words. For example if the definite article "*the*" is mentioned 15 times in a text, it is

only calculated as one different word. Unlike the 'uncorrected' type-token ratio (TTR), which has "been criticized for its sensitivity to sample size, as the ratio tends to decrease as the size of the sample increases" (Lu, 2012, p. 5), the corrected type-token ratio takes into consideration text length. This measure has been repeatedly used by task complexity studies (Ishikawa, 2007; Kuiken and Vedder, 2008; Shajeri and Izadpanah, 2016).

In a recent publication by Crossley et al. (2018) lexical sophistication was looked at as a "multidimensional phenomenon" that consists of various lexical dimensions (features), which together were effective in measuring writing quality, lexical proficiency, and language development. However, because our research does not consider language development between levels, we narrowed down our focus to the dimensions that were found statistically significant in both writing quality and lexical proficiency. The authors looked at Word frequency, Word range, Contextual distinctiveness, Psycholinguistic norms, Word neighborhood, Word recognition, Age of exposure, Semantic relations, N-gram frequency, range, and proportion, N-gram association strength, Academic language. Of the 11 dimensions they tested, seven were found significant in predicting writing quality, and five were significant in measuring lexical proficiency. Given that the latter five also make up five of the previous seven, we ended up using these five measures that showed statistical significance in both writing quality and lexical proficiency. These are n-gram proportions (bi-gram and tri-gram), familiarity, concreteness, AoA (Age of Acquisition), and word specificity.

### E. Fluency Measure

The length of the text was used as a measure of fluency (Larsen-Freeman, 1978; Wolfe-Quintero et. al., 1998). Text length has been used as such by task complexity studies with

significant results (Ishikawa, 2006) as well as insignificant results (Ruiz-Funes, 2015). This measure is especially fitting for the current investigation because learners had to write in a timed condition.

### F. Accuracy Criteria

Similar to syntactic complexity, before measuring accuracy, errors needed to be counted, and for that we followed different rules depending on the type of error, whether spelling or grammatical.

#### 1. Spelling Errors

First, splitting up a compound word was counted as an error. If a learner combined an indefinite article with a misspelled word, two errors were recorded: one for joining the article and one for the misspelling. Deviations from prescriptive capitalization and punctuation rules were not counted as errors. Variations of Latinized proper names and names of local places were not considered errors. For example, "Abdalziz" and "Abdulaziz" were both considered correct. "AL-Slam mall" and "Al-SAlAm mall" were also considered correct.

#### 2. Grammatical Errors

As for the rules governing how we identified grammatical errors, they are as follows. If a sentence could be understood to have either missed two function words or only one, we took the interpretation with fewer errors to be recorded. For example, in the sentence " He was live in Dubai in 2012," we can either say that the learner meant to say: He used to live in Dubai in 2012 or He was living in Dubai in 2012; the latter interpretation would only count one error point in

the example; therefore, we chose it instead of the former interpretation, which would record two error points. The rationale for this criterion is that longer structures tend to be more complex, and learners at the low and intermediate levels will unlikely use such complex structures overall; it is also evident throughout the texts that such complex and long structures are rarely used. Furthermore, sentences with weak structures such as " In my life, I Love some peaple and I do not Love peaple" were not considered erroneous.

### G. Accuracy Measures

For the current study, the definition of accuracy was adopted from Wolfe-Quintero et al., (1998). They defined accuracy as "the ability to be free from errors while using language to communicate'' (p. 33). As for the measures used for detecting errors, we used three measures. The error-free clause ratio (EFCR) was recommended by Evans et al (2014), who compared it to two other measures that were hypothesized to be similar to it (the error-free T-unit ratio & weighted clause ratio). Evans et al (2014) found that no measure "distinguish(ed) these (proficiency) levels as thoroughly as the EFCR." Even before, Evans et al. (2010) quote Wigglesworth (2008) who "reported error-free clauses (EFC) to be the most precise measure of writing accuracy that we currently have." Polio and Shea, (2014) also quote Evans et al. (2010), who used the FACET analysis to check the reliability of a few accuracy measures, including the EFCR, and have found results that indicate "excellent reliability" in the measures. As for task complexity studies, Ishikawa (2006) used the EFCR and has found significant results. Next, the errors per words measure was also used by Ishikawa (2006). The errors per T-unit measure was used by Kuiken and Veddar (2008) to distinguish task complexity effects and was found to be statistically significant.

# IX. RESULTS

## A. Low Proficiency Learners:

In order to answer research question 1, which sought to determine the effects of task complexity on low proficiency learners' written performance measured through accuracy, fluency and complexity (syntactic & lexical), we used 16 measures. Means and standard deviations of each measure are presented in TABLE IV.

As displayed in the table, we used four measures for syntactic complexity, which are the T-units per sentence measure, the Clauses per t-unit measure, and the Clauses per sentence measure, and the Words per clause measure.

Next, we looked into lexical complexity using seven measures. These are the Corrected Type-Token Ratio (CTTR), the Bigram Proportion, the Trigram Proportion, the Familiarity measure, the Concreteness measure, the Age of Acquisition (AoA) measure, and the Word specificity measure.

Third, we used four measures for accuracy. These are the Errors per clause measure, the Errors per t-unit measure, the Error-free clause ratio measure and the Errors per words measure.

Finally, to measure fluency, we used only one measure, which is Word count (W).

**TABLE IV**

DESCRIPTIVE STATISTICS OF CAF MEASURES FOR LOW PROFICIENCY LEARNERS FOR THE SIMPLE AND COMPLEX TASKS

| | | Task | N[c] | Mean | Std. Deviation |
|---|---|---|---|---|---|
| **SYNTACTIC** | **T-units / sentence** | 1.00 [a] | 57.00 | 1.17 | 0.41 |
| | | 2.00 [b] | 15.00 | 1.20 | 0.39 |

| | | | | | |
|---|---|---|---|---|---|
| **COMPLEXITY MEASURES** | **Clause / t-unit** | 1.00 [a] | 57.00 | 1.09 | 0.04 |
| | | 2.00 [b] | 15.00 | 1.14 | 0.07 |
| | **Clause / Sentence** | 1.00 [a] | 57.00 | 1.27 | 0.44 |
| | | 2.00 [b] | 15.00 | 1.35 | 0.40 |
| | **Words / Clause** | 1.00 [a] | 57.00 | 7.18 | 1.89 |
| | | 2.00 [b] | 15.00 | 6.51 | 1.14 |
| **LEXICAL COMPLEXITY MEASURES** | **CTTR** | 1.00 [a] | 57.00 | 0.14 | 0.02 |
| | | 2.00 [b] | 15.00 | 0.15 | 0.02 |
| | **Bigram Proportion** | 1.00 [a] | 57.00 | 0.51 | 0.09 |
| | | 2.00 [b] | 15.00 | 0.52 | 0.13 |
| | **Trigram Proportion** | 1.00 [a] | [b] 57.00 | 0.13 | 0.06 |
| | | 2.00 [b] | 15.00 | 0.12 | 0.06 |
| | **Familiarity** | 1.00 [a] | 57.00 | 599.32 | 5.47 |
| | | 2.00 [b] | 15.00 | 593.00 | 11.09 |
| | **Concreteness** | 1.00 [a] | 57.00 | 419.50 | 38.75 |
| | | 2.00 [b] | 15.00 | 397.49 | 26.37 |
| | **Age of Acquisition** | 1.00 [a] | 57.00 | 4.50 | 0.28 |
| | | 2.00 [b] | 15.00 | 4.73 | 0.33 |
| | **Word Specificity** | 1.00 [a] | 57.00 | 2.36 | 0.16 |
| | | 2.00 [b] | 15.00 | 2.51 | 0.18 |
| **ACCURACY** | **Errors / Clauses** | 1.00 [a] | 57.00 | 1.51 | 0.87 |
| | | 2.00 [b] | 15.00 | 1.39 | 0.60 |
| | **Errors / T-units** | 1.00 [a] | 57.00 | 1.66 | 0.99 |
| | | 2.00 [b] | 15.00 | 1.60 | 0.79 |
| | **EFCR** | 1.00 [a] | 57.00 | 0.31 | 0.24 |
| | | 2.00 [b] | 15.00 | 0.23 | 0.16 |
| | **Errors / Words** | 1.00 [a] | 57.00 | 0.22 | 0.12 |
| | | 2.00 [b] | 15.00 | 0.21 | 0.08 |
| **FLUENCY** | **W** | 1.00 [a] | 57.00 | 64.58 | 28.05 |
| | | 2.00 [b] | 15.00 | 76.33 | 32.00 |

[a] 1= simple task.

[b] 2= complex task.

[c] N= sample number

Before conducting inferential statistics to determine statistical significance of the resulting differences in means, we needed to account for the assumptions of homogeneity and normality.

So, first, the data of all the measures in the simple and complex tasks were submitted to the Shapiro-Wilk test of normality. The results, as displayed in TABLE V, show that only six of the sixteen measures failed the null hypothesis and so are normally-distributed (> 0.05). These are the CTTR measure, the Bigram proportion, the Trigram proportion, the Concreteness measure, the AoA measure, and the Word Specificity measure; these measures all assessed lexical complexity.

**TABLE V**

TEST OF NORMALITY OF SIMPLE TASK AND COMPLEX TASK IN LOW PROFICIENCY LEARNERS

| | **Shapiro-Wilk** | | |
|---|---|---|---|
| | Statistic | df [a] | Sig. [b] |
| **T-units / sentence** | .745 | 72 | .000* |
| **Clause / t-unit** | .871 | 72 | .000* |

| | | | |
|---|---|---|---|
| **Clause / Sentence** | .697 | 72 | .000* |
| **Words / Clause** | .932 | 72 | .001* |
| **CTTR** | .994 | 72 | .984 |
| **Bigram Proportion** | .981 | 72 | .357 |
| **Trigram Proportion** | .982 | 72 | .401 |
| **Familiarity** | .902 | 72 | .000* |
| **Concreteness** | .989 | 72 | .765 |
| **Age of Acquisition** | .969 | 72 | .070 |
| **Word Specificity** | .992 | 72 | .946 |
| **Errors / Clauses** | .954 | 72 | .011* |
| **Errors / T-units** | .943 | 72 | .003* |
| **EFCR** | .921 | 72 | .000* |
| **Errors / Words** | .966 | 72 | .049* |
| **W** | .887 | 72 | .000* |

[a] df= degrees of freedom

[b] Sig.= significance

Second, Levene's test for homogeneity was used on these six normally-distributed measures. Results in TABLE VI show that the null hypothesis has also failed for all the measures; therefore, they are all variance-homogenous, which in turn mean that the T-test is the appropriate statistical test for these six measures.

**TABLE VI**

TEST OF HOMOGENEITY OF SIMPLE TASK AND COMPLEX TASK IN LOW PROFICIENCY LEARNERS

| | **Levene's Test for Equality of Variances** | |
|---|---|---|
| | *F* | *Sig.* [a] |
| **CTTR** | .250 | .62 |
| **Bigram proportion** | 3.99 | .05 |
| **Trigram proportion** | .021 | .89 |
| **AoA** | .603 | .44 |
| **Word specificity** | 0.1 | .75 |
| **Err/W** | 2.34 | 1.3 |

[a] Sig.= significance

Since the tests for the homogeneity and normality revealed that six measures are both homogenous and normally-distributed, we ran the Independent-Samples T-test. As shown in TABLE VII, out of the six measures, Word specificity, Errors per words, and Word count show statistically significant differences between the simple task and the complex task. The Word specificity measure showed that words with specific meanings were statistically used more by learners in the complex task (M = 2.51, SD = 0.18) than by learners in the simple task (M= 2.36, SD = 0.16), t = 2.07, p = 0.04, CI = [-0.25, -0.06]. The Errors per words measure showed that learners in the complex task used less errors per total number of words (M = 0.21, SD = 0.08) than the simple task (M = 0.22, SD = 0.12), t = -2.76, p = 0.01, CI = [-0.06, 0.07]. Lastly, the Word count revealed that learner in the complex task elicited significantly longer texts (M = 76.33, SD =32) than learners in the simple task (M= 64.58, SD = 28.05), t = -3.21, p = 0.00, CI = [-28.47, 4.96].

**TABLE VII**

T-TEST COMPARING SIMPLE AND COMPLEX TASK MEASURES IN LOW PROFICIENCY LEARNERS

| | **T-test for Equality of Means** | | | | | | |
|---|---|---|---|---|---|---|---|
| | *Simple task Mean (SD)* | *Complex task Mean (SD)* | *t* [a] | *df* | *Sig. (2-tailed)* [b] | *95% CI for Mean Difference* [c] | |
| | | | | | | *Lower bound* | *Upper bound* |
| **CTTR** | 0.14 (0.02) | 0.15 (0.02) | -1.01 | 70 | 0.32 | -0.02 | 0.01 |
| **Trigram proportion** | .13 (0.06) | .12 (0.06) | -0.52 | 70 | 0.61 | -0.03 | 0.04 |
| **AoA** | 4.5 (0.28) | 4.7 (0.33) | 0.38 | 70 | 0.71 | -0.40 | -0.06 |
| **Word specificity** | 2.36 (0.16) | 4.73 (0.18) | 2.07 | 70 | 0.04* | -0.25 | -0.06 |
| **Err/W** | 0.22 (0.12) | .21 (0.08) | -2.76 | 70 | 0.01* | -0.06 | 0.07 |
| **W** | 64.58 (28.05) | 76.33 (32.00) | -3.21 | 70 | 0.00* | -28.47 | 4.96 |

[a] t= t-value

[b] Sig.= significance

[c] CI= confidence intervals

As for the measures with data that were not normally-distributed, they were submitted to the non-parametric Mann-Whitney U test. The results are presented in TABLE VIII for the mean ranks and in TABLE IX for statistical significance.

**TABLE VIII**

MEAN RANKS OF SIMPLE TASK AND COMPLEX TASK IN LOW PROFICIENCY LEARNERS

| **Ranks** | | | | |
|---|---|---|---|---|
| | *Task* | *N* [c] | *Mean Rank* | *Sum of Ranks* |
| **T-units / sentence** | 1 | 57 | 35.58 | 2028.00 |
| | 2 | 15 | 40.00 | 600.00 |
| | Total | 72 | | |
| **Clause / t-unit** | 1 | 57 | 33.19 | 1892.00 |
| | 2 | 15 | 49.07 | 736.00 |
| | Total | 72 | | |
| **Clause / Sentence** | 1 [a] | 57 | 34.75 | 1980.50 |
| | 2 [b] | 15 | 43.17 | 647.50 |
| | Total | 72 | | |
| **Words / Clause** | 1 [a] | 57 | 38.02 | 2167.00 |
| | 2 [b] | 15 | 30.73 | 461.00 |
| | Total | 72 | | |
| **Bigram Proportion** | 1 [a] | 57 | 35.67 | 2033.00 |

| | | | | |
|---|---|---|---|---|
| | 2 [b] | 15 | 39.67 | 595.00 |
| | Total | 72 | | |
| **Familiarity** | 1 [a] | 57 | 39.21 | 2235.00 |
| | 2 [b] | 15 | 26.20 | 393.00 |
| | Total | 72 | | |
| **Concreteness** | 1 [a] | 57 | 39.25 | 2237.00 |
| | 2 [b] | 15 | 26.07 | 391.00 |
| | Total | 72 | | |
| **Errors / Clauses** | 1 [a] | 57 | 36.75 | 2094.50 |
| | 2 [b] | 15 | 35.57 | 533.50 |
| | Total | 72 | | |
| **Errors / T-units** | 1 [a] | 57 | 36.51 | 2081.00 |
| | 2 [b] | 15 | 36.47 | 547.00 |
| | Total | 72 | | |
| **EFCR** | 1 [a] | 57 | 37.68 | 2147.50 |
| | 2 [b] | 15 | 32.03 | 480.50 |
| | Total | 72 | | |

[a] 1= simple task

[b] 2= complex task

[c] N= sample number.

**TABLE IX**

SIGNIFICANCE OF MEAN RANKS OF SIMPLE TASK AND COMPLEX TASK IN LOW PROFICIENCY LEARNERS

| | **Mann-Whitney U** | **Z** [a] | **Asymp. Sig. (2-tailed)** [b] |
|---|---|---|---|
| **T-units/sentence** | 375 | -.728 | .466 |
| **Clause/t-unit** | 239 | -2.622 | .009* |
| **Clause/Sentence** | 327.5 | -1.429 | .153 |
| **Words/Clause** | 341 | -1.200 | .230 |
| **Bigram Proportion** | 380 | -.659 | .510 |
| **Familiarity** | 273 | -2.142 | .032* |
| **Concreteness** | 271 | -2.17 | .030* |
| **Errors/Clauses** | 413.5 | -.194 | .846 |
| **Errors/T-units** | 427 | -.007 | .994 |
| **EFCR** | 360.5 | -.932 | .351 |

[a] Z=z-value

[b] Sig.= significance.

First, looking at the syntactic complexity measure Clause per t-unit (subordination), we see that learners in the complex task had more subordination (rank m = 49.07) than did learners in the simple task (rank m = 33.19), U= 239, p = 0.01, r = 0.31.

The other two measures were for lexical complexity. Beginning with Familiarity, the Mann Whitney test showed that learners used less familiar words (more complex) (rank m = 26.20) in the complex task as compared to learners in the simple task (rank m = 39.21), U = 273,

p = 0.03, r = 0.25. Second and last, the Concreteness measure showed that learners in the complex task used less concrete words (more complex) (rank m = 26.07) than did learners in the simple task (rank r = 39.25), U = 271, p = 0.03, r = 0.26.

**TABLE X**

SUMMARY OF RESULTS

| **T-test** | **M (SD) 1** | **M (SD) 2** | **t** [a] | **p** [b] | **CI** [c] | **g** [d] |
|---|---|---|---|---|---|---|
| **Word specificity** | 2.36 (0.16) | 2.51 (0.18) | 2.07 | 0.04 | [-0.25, -0.06] | 0.91 |
| **Err/w** | 0.22 (0.08) | 0.21 (0.12) | -2.76 | 0.01 | [-0.06, 0.07] | 0.11 |
| **W** | 64.58 (28.05) | 76.33 (32) | -3.21 | 0.00 | [-28.47, 4.96] | 0.41 |
| **Mann Whitney** | **Rank m 1** | **Rank m 2** | **U** [e] | **p** [b] | **r** [f] | |
| **Clauses/t-unit** | 33.19 | 49.07 | 239 | 0.01 | 0.31 | |
| **Familiarity** | 39.21 | 26.20 | 273 | 0.03 | 0.25 | |
| **Concreteness** | 39.25 | 26.07 | 271 | 0.03 | 0.26 | |

[a] t= t-value.

[b] p= p-value.

[c] CI= confidence intervals.

[d] g= Hedge's g for effect size.

[e] U= Mann Whitney's U.

[f] r= effect size.

Looking at the results in their entirety, we conclude that out of the 16 measures we used in the comparison between the complex task and the simple task for the low proficiency learners group, only six measures showed statistical significance. These are Clause per t-unit for syntactic complexity, Word specificity, Familiarity, and Concreteness for lexical complexity, Errors per words for accuracy, and Word count for fluency (TABLE X).

The results show that learners in the complex task elicited more complex syntax than learner in the simple task by resulting in more subordinate clauses with the small effect size of 0.31. For lexical complexity, learners in the complex task elicited more complex lexical items in the Familiarity and Concreteness measures by producing less familiar and concrete words with the small effect sizes of 0.25 and 0.26, respectively. learners in the complex task also used more lexically complex items by producing words with specific meanings with the large effect size of 0.88. As for accuracy, learners in the complex task produced more accurate production measured by fewer errors per words, however the effect size was too small to matter (0.1). Finally, learners in the complex task used more words on average than did learners in the simple task with a medium effect size of 0.39.

To summarize, in the comparison between learners in the complex task and learners in the simple task in the low proficiency level, the complex task elicited more complex (syntactic & lexical) and fluent language, while no statistically significant differences were found in accuracy.

**B. Intermediate Proficiency Learners**

As with question 1, we followed the same statistical procedures in question 2. We first present the descriptive statistics of the two tasks, followed by a test of normality. Data that are normally distributed are submitted to a homogeneity test. If it proves to be both normally distributed and homogeneous, we run a t-test to for statistical significance. If it fails the normality test, it is submitted to the nonparametric test, Mann-Whitney U.

First, means and standard deviations of the sixteen measures of complexity (syntactic & lexical), accuracy and fluency are presented in TABLE XI.

**TABLE XI**

DESCRIPTIVE STATISTICS OF ACCURACY, COMPLEXITY, AND FLUENCY MEASURES FOR LOW PROFICIENCY LEARNERS FOR THE SIMPLE AND COMPLEX TASKS

| | | Task | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| **SYNTACTIC COMPLEXITY MEASURES** | **T-units / sentence** | 1.00 [a] | 29.00 | 1.28 | 0.31 |
| | | 2.00 [b] | 26.00 | 1.29 | 0.88 |
| | **Clause / t-unit** | 1.00 [a] | 29.00 | 1.19 | 0.05 |
| | | 2.00 [b] | 26.00 | 1.42 | 1.00 |
| | **Clause / Sentence** | 1.00 [a] | 29.00 | 1.51 | 0.35 |
| | | 2.00 [b] | 26.00 | 1.92 | 1.10 |
| | **Words / Clause** | 1.00 [a] | 29.00 | 7.55 | 1.28 |
| | | 2.00 [b] | 26.00 | 13.83 | 4.75 |
| **LEXICAL COMPLEXITY** | **CTTR** | 1.00 [a] | 29.00 | 0.12 | 0.02 |
| | | 2.00 [b] | 26.00 | 0.12 | 0.02 |

| | | | | | |
|---|---|---|---|---|---|
| **MEASURES** | **Bigram Proportion** | 1.00 [a] | 29.00 | 0.50 | 0.11 |
| | | 2.00 [b] | 26.00 | 0.60 | 0.07 |
| | **Trigram Proportion** | 1.00 [a] | 29.00 | 0.11 | 0.06 |
| | | 2.00 [b] | 26.00 | 0.17 | 0.06 |
| | **Familiarity** | 1.00 [a] | 29.00 | 592.94 | 9.31 |
| | | 2.00 [b] | 26.00 | 595.94 | 4.39 |
| | **Concreteness** | 1.00 [a] | 29.00 | 385.19 | 24.83 |
| | | 2.00 [b] | 26.00 | 366.62 | 18.79 |
| | **Age of Acquisition** | 1.00 [a] | 29.00 | 4.98 | 0.47 |
| | | 2.00 [b] | 26.00 | 5.16 | 0.35 |
| | **Word Specificity** | 1.00 [a] | 29.00 | 2.45 | 0.12 |
| | | 2.00 [b] | 26.00 | 2.53 | 0.09 |
| **ACCURACY** | **Errors / Clauses** | 1.00 [a] | 29.00 | 1.02 | 0.60 |
| | | 2.00 [b] | 26.00 | 2.30 | 1.02 |
| | **Errors / T-units** | 1.00 [a] | 29.00 | 1.22 | 0.74 |
| | | 2.00 [b] | 26.00 | 3.11 | 3.37 |
| | **EFCR** | 1.00 [a] | 29.00 | 0.22 | 0.18 |
| | | 2.00 [b] | 26.00 | 0.51 | 0.32 |
| | **Errors / Words** | 1.00 [a] | 29.00 | 0.14 | 0.08 |
| | | 2.00 [b] | 26.00 | 0.17 | 0.06 |
| **FLUENCY** | **W** | 1.00 [a] | 29.00 | 133.83 | 35.18 |
| | | 2.00 [b] | 26.00 | 155.00 | 42.62 |

[a] 1= simple task.

[b] 2= complex task.

[c] N= sample number.

Starting with the test of normality, the data of all the measures in the simple task and in the complex task were submitted to the Shapiro-Wilk test. Results in TABLE XII show that only six of the sixteen measures failed the null hypothesis and so are normally-distributed (> 0.05). These are the lexical complexity measures of the CTTR, the Trigram proportion, the AoA measure, and the Word specificity, the accuracy measure of Error per words and the fluency measure of Word count.

**TABLE XII**

TEST OF NORMALITY OF SIMPLE TASK AND COMPLEX TASK IN LOW PROFICIENCY LEARNERS

| | **Shapiro-Wilk** | | |
|---|---|---|---|
| | Statistic | df [a] | Sig. [b] |
| **T-units / sentence** | .910 | 55 | .001* |
| **Clause / t-unit** | .473 | 55 | .000* |
| **Clause / Sentence** | .852 | 55 | .000* |
| **Words / Clause** | .818 | 55 | .000* |
| **CTTR** | .988 | 55 | .874 |
| **Bigram Proportion** | .945 | 55 | .014* |
| **Trigram Proportion** | .986 | 55 | .761 |
| **Familiarity** | .926 | 55 | .002* |
| **Concreteness** | .950 | 55 | .022* |
| **Age of Acquisition** | .980 | 55 | .493 |

| | | | |
|---|---|---|---|
| **Word Specificity** | .965 | 55 | .112 |
| **Errors / Clauses** | .914 | 55 | .001* |
| **Errors / T-units** | .769 | 55 | .000* |
| **EFCR** | .871 | 55 | .000* |
| **Errors / Words** | .975 | 55 | .305 |
| **W** | .983 | 55 | .624 |

[a] df= degrees of freedom.

[b] Sig.= significance.

Second, we ran Levene's test for homogeneity. Results are shown in TABLE XIII, and they indicate that the six normally-distributed measures were also variance-homogenous. Therefore, by accounting for both statistical assumptions (normality & homogeneity), the T-test is the appropriate statistical test.

**TABLE XIII**

TEST OF HOMOGENEITY OF SIMPLE TASK AND COMPLEX TASK IN LOW PROFICIENCY LEARNERS

| | **Levene's Test for Equality of Variances** | |
|---|---|---|
| | *F* | *Sig.* [a] |
| **CTTR** | .578 | .450 |

| | | |
|---|---|---|
| **Trigram** | .010 | .919 |
| **AoA** | 1.486 | .228 |
| **Word sp** | 1.486 | .228 |
| **Err/W** | 1.159 | .287 |
| **W** | .592 | .445 |

[a] Sig.= significance

Consequently, we ran the Independent-Samples T-test on these data and the results are displayed in TABLE XIV. They show that statistically significant differences were found in three of the six included measures, which are the lexical complexity measures of the Trigram proportion and Word specificity, and the fluency measure of Word count. Starting with the Trigram proportion, it shows that learners in the complex task used more contiguous sequences of *n* words, such as *across the sky* and *and he said* (M = 0.17, SD = 0.06) than did learners in the simple task (M= 0.11, SD = 0.06), t = -3.47, p = 0.00, CI = [-0.09, -0.02]. The Word specificity measure shows that learners in the complex task also used significantly more words with specific meanings (M = 2.53, SD = 0.09) than did learners in the simple task (M = 2.45, SD = 0.12), t = -1.55, p = 0.00, CI = [-0.14, -0.03]. Lastly, the Word count revealed that learners in the complex task produced significantly longer texts (M = 155, SD =42.62) than did learners in the simple task (M = 133.83, SD =35.18), t = -2.02, p = .049, CI = [-42.23, -0.12].

**TABLE XIV**

T-TEST COMPARING SIMPLE AND COMPLEX TASK MEASURES IN LOW PROFICIENCY LEARNERS

| | **T-test for Equality of Means** | | | | | | |
|---|---|---|---|---|---|---|---|
| | *Simple Mean (SD)* | *Complex Mean (SD)* | *t* [a] | *df* | *Sig.(2-tailed)*[b] | *95% CI for Mean Difference* [c] | |
| | | | | | | *Lower bound* | *Upper bound* |
| **CTTR** | 0.12 (0.02) | 0.12 (0.02) | .339 | 53 | .736 | -0.01 | 0.01 |
| **Trigram proportion** | 0.11 (0.06) | 0.17 (0.06) | -3.469 | 53 | .001* | -0.09 | -0.02 |
| **AoA** | 4.98 (0.47) | 5.16 (0.35) | -1.545 | 53 | .128 | -0.40 | 0.05 |
| **Word specificity** | 2.45 (0.12) | 2.53 (0.09) | -1.545 | 53 | .003* | -0.14 | -0.03 |
| **Err/W** | 0.14 (0.08) | 0.17 (0.06) | -1.836 | 53 | .072 | -0.07 | 0.00 |
| **W** | 133.83 (35.18) | 155 (42.62) | -2.017 | 53 | .049* | -42.23 | -0.12 |

[a] t= t-value.

[b] Sig.= significance.

[c] CI= confidence intervals.

As for the data that were not normally-distributed, they were submitted to the nonparametric Mann-Whitney U test. The mean ranks are presented in Table XV, and the test for statistical significance is presented in Table XVI.

**TABLE XV**

MEAN RANKS OF SIMPLE TASK AND COMPLEX TASK IN LOW PROFICIENCY LEARNERS

| **Ranks** | | | | |
|---|---|---|---|---|
| | *Task* | *N* [c] | *Mean Rank* | *Sum of Ranks* |
| **T-units / sentence** | 1 [a] | 29 | 28.62 | 830.00 |
| | 2 [b] | 26 | 27.31 | 710.00 |
| | Total | 55 | | |
| **Clause / t-unit** | 1 [a] | 29 | 16.14 | 468.00 |
| | 2 [b] | 26 | 41.23 | 1072.00 |
| | Total | 55 | | |
| **Clause / Sentence** | 1 [a] | 29 | 24.28 | 704.00 |
| | 2 [b] | 26 | 32.15 | 836.00 |
| | Total | 55 | | |
| **Words / Clause** | 1 [a] | 29 | 15.93 | 462.00 |
| | 2 [b] | 26 | 41.46 | 1078.00 |
| | Total | 55 | | |
| **Bigram Proportion** | 1 [a] | 29 | 36.12 | 1047.50 |
| | 2 [b] | 26 | 18.94 | 492.50 |

| | | | | |
|---|---|---|---|---|
| | Total | 55 | | |
| **Familiarity** | 1 [a] | 29 | 20.38 | 591.00 |
| | 2 [b] | 26 | 36.50 | 949.00 |
| | Total | 55 | | |
| **Concreteness** | 1 [a] | 29 | 21.45 | 622.00 |
| | 2 [b] | 26 | 35.31 | 918.00 |
| | Total | 55 | | |
| **Errors / Clauses** | 1 [a] | 29 | 25.88 | 750.50 |
| | 2 [b] | 26 | 30.37 | 789.50 |
| | Total | 55 | | |
| **Errors / T-units** | 1 [a] | 29 | 33.55 | 973.00 |
| | 2 [b] | 26 | 21.81 | 567.00 |
| | Total | 55 | | |
| **EFCR** | 1 [a] | 29 | 24.34 | 706.00 |
| | 2 [b] | 26 | 32.08 | 834.00 |
| | Total | 55 | | |

[a] 1= simple task.

[b] 2= complex task.

[c] N= sample number.

**TABLE XVI**

SIGNIFICANCE OF MEAN RANKS OF SIMPLE TASK AND COMPLEX TASK IN LOW PROFICIENCY LEARNERS

| | **Mann-Whitney U** | **Z** [a] | **Asymp. Sig. (2-tailed)** [b] |
|---|---|---|---|
| **T-units / sentence** | 359.000 | -.303 | .762 |
| **Clause / t-unit** | 33.000 | -5.806 | .000* |
| **Clause / Sentence** | 269.000 | -1.821 | .069 |
| **Words / Clause** | 27.000 | -5.900 | .000* |
| **Bigram Proportion** | 156.000 | -3.726 | .000* |
| **Familiarity** | 315.500 | -1.037 | .300 |
| **Concreteness** | 216.000 | -2.714 | .007* |
| **Errors / Clause** | 88.000 | -4.873 | .000* |
| **Errors / T-unit** | 74.000 | -5.108 | .000* |
| **EFCR** | 147.000 | -3.878 | .000* |

[a] Z= z-value.

[b] Sig.= significance.

Looking first at the syntactic complexity measures, the Mann-Whitney test indicated that the Clause per t-unit (subordination) was greater for the complex task (rank m = 41.23) than for the simple task (rank m = 16.14), U= 33, p = 0.0, r = 0.78. The other syntactic measure, Words per clause, shows that learners in the complex task used significantly longer clauses (rank m = 41.46) than did learners in the simple task (rank m = 15.93) U= 27, p = 0.0, r = 0.8.

For the analysis of lexical complexity, results of the Bigram proportion shows that learners in the complex task had statistically less contiguous sequences of n words, such as *in the* or *to make* (rank m = 18.94), than did the learners in the simple task (rank m = 36.12) U= 156, p = 0.0, r = 0.5. The lexical complexity measure, Concreteness, also reveals that learners in the complex task had more concrete words (less complex) (rank m = 35.31) than did learners in the simple task (rank m = 21.45), U= 216, p = 0.0, r = 0.37.

The last three measures concern accuracy. The Errors per clause measure reveals that learners in the complex task had more errors per clause (rank m = 30.37) than did learners in the simple task (rank m = 25.88), U = 88, p = 0.01, r = 0.66. The Errors per t-unit measure shows that learners in the complex task had less errors per t-unit (rank m = 21.81) than did learners in the simple task (rank m = 33.55), U = 74, p = 0.0, r = 0.69. Lastly, the EFCR measure shows that learners in the complex task had more error-free clauses (rank m = 32.08) than did learners in the simple task (rank m = 24.34), U = 147, p = 0.0, r = 0.52.

**XVII**

SUMMARY OF RESULTS

| **T-test** | **M (SD) 1** | **M (SD) 2** | **t** [a] | **p** [b] | **CI** [c] | **g** [d] |
|---|---|---|---|---|---|---|
| **Trigram proportion** | 0.11 (0.06) | 0.17 (0.06) | -3.47 | 0.00 | [-0.09, -0.02] | 1 |
| **Word specificity** | 2.45 (0.12) | 2.53 (0.09) | -1.55 | 0.00 | [-0.14, -0.03] | 0.75 |
| **W** | 133.83 (35.18) | 155 (42.62) | -2.02 | 0.49 | [-42.23, -0.12] | 0.54 |

| | | | | | |
|---|---|---|---|---|---|
| **Mann Whitney** | **Rank m 1** | **Rank m 2** | **U** [e] | **p** [b] | **r** [f] |
| **Clauses/t-unit** | 16.14 | 41.23 | 33 | 0.00 | 0.78 |
| **Words/clause** | 15.93 | 41.46 | 27 | 0.00 | 0.8 |
| **Bigram proportion** | 36.12 | 18.94 | 156 | 0.00 | 0.5 |
| **Concreteness** | 21.45 | 35.31 | 216 | 0.00 | 0.37 |
| **Err/C** | 25.88 | 30.37 | 88 | 0.01 | 0.66 |
| **Err/T** | 33.55 | 21.81 | 74 | 0.00 | 0.69 |
| **EFCR** | 24.34 | 32.08 | 147 | 0.00 | 0.52 |

[a] t= t-value.

[b] p= p-value.

[c] CI= Confidence intervals.

[d] g= hedge's g for effect size.

[e] U= Mann Whitney's U.

[f] r= effect size.

Looking at the results in their entirety, we conclude that out of the 16 measures we used in the comparison between the complex task and the simple task for the intermediate proficiency learners group, ten measures showed statistical significance. These are the Clauses per t-unit and Words per clause for syntactic complexity, the Bigram proportion, Trigram proportion,

Concreteness and Word specificity for lexical complexity, the Errors per clause, Errors per T-unit and EFCR for accuracy, and Word count for accuracy (TABLE XIX).

The results show that learners in the complex task produced more complex syntax than learners in the simple task resulting in more subordinate clauses and longer clauses overall with the large effect sizes of 0.78 and 0.8, respectively. As for lexical complexity, learners in the complex task used less complex lexical items in the Bigram proportion and Concreteness measures by producing less native-like Bigrams and more concrete words with the medium effect sizes of 0.5 and 0.37, respectively. However, the other two lexical measures, Trigram proportion and Word specificity, revealed the opposite results with more lexical complexity in the complex task with the large effect sizes of 1.00 and 0.75, respectively. In order to resolve this apparent contradiction, we weighted the measures in terms of complexity. Beginning with the Bigrams and Trigrams put opposite to each other, we argue that trigrams are more indicative of complexity because they look at the sequence of three words that are native-like rather than only two. We then compared the Word specificity measure and the Concreteness measure because they both look at the semantic aspect of the lexical items. We argue that unlike Concreteness, the Word specificity measure showed statistical significance in the first research question. Therefore, Word specificity was a stronger indicator of lexical complexity, especially because that it was the only measure with a large effect size (0.91) compared to all the CAF measures in the low proficiency group. Lastly, the effect sizes of the Bigram proportion and the Concreteness measures are medium, while they are large for the Trigram proportion and Word specificity. For accuracy, learners in the complex task used more errors per clause with a medium effect size of 0.66. However, the Errors per t-unit and Error-free clause ratio measures show the opposite results: more accuracy by learners in the complex task with the effects sizes of 0.69 and 0.52,

respectively. To resolve this contradiction, we judged Errors per clause measure inappropriate in this case because clauses in the complex task were both more in number and in length as we saw with syntactic complexity at the beginning of the paragraph. So, in order to measure accuracy accurately, the measure used has to be based on an aspect of the production that is equal in both data sets (e.g. unlike here - *clauses*). Finally, the complex task elicited more words on average than the simple task with the medium effect size of 0.54.

In summary, in the comparison between learners in the complex task and learners in the simple task in the low proficiency level, the complex task elicited more complex (syntactic & lexical), fluent and accurate language.

# X. DISCUSSION

## A. Low Proficiency Learners

As we saw, when testing the effects of task complexity on the written production of low proficiency level learners, we saw that learners in the complex task used more complex (syntactic & lexical) and fluent language, while no differences were found in accuracy.

These results are not in line with Skehan's Trade-Off Hypothesis. Skehan (1996; 1998; 2014; Skehan & Foster, 1999; 2001) claims that a trade-off effect between the three constructs of performance (complexity, accuracy, and fluency) will occur, namely that some constructs will increase at the expense of other constructs in the complex task. And so, because the constructs of both complexity (syntactic & lexical) and fluency increased in the complex task while accuracy did *not* decrease as compared to the simple task, the results are not in line with Skehan's hypothesis, and a trade-off effect did not occur.

On the other hand, these results are *partially* in line with Robinson's Cognition Hypothesis. Robinson claims (1996a; 2001; 2003a; 2011a; 2011b; Baralt et. al., 2014; Robinson & Gilabert, 2007a) that raising task complexity in the resource-directing variables in his Triadic Componential Framework will elicit greater complexity and accuracy from learners. Because the complex task only elicited greater complexity, while accuracy was maintained equal to the simple task, we concluded that the results are *partially* in line with Robinson's hypothesis.

Looking at previous literature, we find that our results are similar to Hosseini and Rahimpour (2010) and Salimi et al., (2011). These two researchers found a significant increase in syntactic complexity with no increase nor decrease in the level of accuracy

**B. Intermediate Proficiency Learners**

When testing the effects of task complexity on the written production of the intermediate proficiency level learners, we saw that learners in the complex task used more complex (syntactic & lexical), fluent, and accurate language.

As Skehan (1996; 1998; 2014; Skehan & Foster, 1999; 2001) predicts the occurrence of a trade-off, the results contradict his viewpoint because all the constructs of performance increased in the complex task in our study. On the other hand, these results are in full support of Robinson's hypothesis. He predicts task complexity will increase complexity and accuracy, and this is what happened to our study.

Looking at previous literature, we found that our results partially support Ishikawa (2007). He found that syntactic complexity, accuracy and fluency all increased in the complex task, but effects were found in lexical complexity. In our study, lexical complexity was also higher in the complex task.

**C. Task Complexity and Proficiency Level Comparison:**

In this section, we will mention three observations that were noticed when comparing the results of the two proficiency levels. First, although the complex tasks in both proficiency levels elicited higher syntactic and lexical complexity as well as fluency, it only elicited higher accuracy in the intermediate proficiency group, whereas no accuracy effects were found in the low proficiency level group. Second, unlike the intermediate proficiency group, which showed that 10 out of the 16 CAF measures revealed statistically significant differences, the low proficiency group showed statistical significance in only 6 measures. Lastly, the effect sizes of these 6 measures in the low proficiency learners were also low overall except for one: 0.11, 0.25,

0.26, 0.31, 0.41, 0.91 (see TABLE X). One the other hand, the effect sizes of the differences between the tasks in the intermediate proficiency level were either medium or large: 0.37, 0.5, 0.52, 0.54, 0.66, 0.69, 0.75, 0.78, 0.8, 1.00 (see TABLE XVII). In other words, the complex task in the intermediate proficiency group, as compared to the complex task in the low proficiency group, showed increases in all the constructs (complexity (syntactic & lexical), accuracy and fluency); these increases were found in more measures (10 vs. 6); and they were larger (mostly low vs. medium to high). The possible explanations to these observations will be discussed in the next section.

## XI. LIMITATIONS

In this section, three limitations to the study design will be discussed as probable explanations to the overall results and to task complexity/proficiency level interaction. First, the complexifying resource-directing variables (according to Robinson's framework), were different in the two proficiency groups. The low proficiency group had −/+ causal reasoning as its complexifying element, whereas the intermediate proficiency group had the +/− here-and-now. Therefore, using different complexifying elements in the two proficiency groups has added another variable that complicates the interpretation of the results. Accordingly, in future studies, we recommend using the same variable in both proficiency groups. Another limitation which we predict might have skewed the results in the low proficiency group is that the complex task assigned to the low proficiency learners was not significantly more complex than the simple task. The complex task (Appendix A) asked students to write about an important person in their lives and provided learners with a structure and steps to be followed: **1)** who the person is and where the person is from; **2)** What the person does and where the person works and studies; **3)** Why the person is important to you and why you like him/her. We categorized this task as complex because the third step of the task asked students to explain the reasoning behind choosing a particular person (+causal reasoning). Thus, only one third of the complex task was complex. On the other hand, the complex task in the intermediate proficiency group was completely about events that will happen in the future (there-and-then) (Appendix A). Lastly, the number of participants in all four groups (simple/complex task in low proficiency learners - simple/complex task in intermediate proficiency learners) was not equal. Whereas the difference between the tasks in the intermediate proficiency group was small (simple task N= 29 - complex task N= 26), the difference was large in the low proficiency group (simple task N= 57 - complex task N= 15).

Consequently, next to the first and second limitation, this could explain the apparent differences between the results of the two proficiency levels.

## XII. CONCLUSION

From our results and the results of those who preceded us in this line of research (Cho, 2015; Hosseini & Rahimpour, 2010; Ishikawa, 2007; Kuiken and Vedder, 2008; Mohammadzadeh Mohammadabadi et al., 2013; Ruiz-Funes's, 2015; Salimi et al., 2011), certain patterns emerge that are beneficial to the overall research and teaching community. First, task complexity in the writing modality seems to never decrease performance (operationalized though the CAF constructs). Complex tasks either increase or maintain the same level of linguistic complexity, accuracy or fluency as compared to the simple task. Second, complex tasks tend to increase linguistic complexity more often (Hosseini & Rahimpour, 2010; Ishikawa, 2007; Salimi et al., 2011) than accuracy (Ishikawa, 2007; Kuiken and Vedder, 2008). Third, fluency is more likely to increase in the complex task (Cho, 2015; Ishikawa, 2007; Kuiken and Vedder, 2008).

## XIII. IMPLICATIONS

In our study design, we only focused on resource-directing variables. Within these, we only looked at the -/+here-and-now and -/+causal reasoning. Both were proposed by Robinson and Gilabert (2007) as variables that direct students' attention to linguistic complexity and accuracy. In this regard, both variables proved to mostly do what Robinson claims they do.

Results of our study indicate that manipulating tasks in the -/+here-and-now resource-directing variable has yielded positive results confirming Robinson's Cognition Hypothesis. These results are characterized by statistically significant increases in all the indicators of performance (accuracy, fluency, syntactic and lexical complexity) in the complex task. We, therefore, recommend using this variable in sequencing tasks complexity.

On the other hand, results of manipulating tasks in the -/+causal reasoning resource-directing variable have only partially confirmed Robinson's, which could be due to limitations in the study design. We, therefore, encourage readers to keep in mind these limitations while interpreting the results of our study. We also encourage future studies to exercise more caution in choosing tasks that follow Robinson's framework manipulations as accurately as possible.

## CITED LITERATURE

Baralt, M., Gilabert, R., & Robinson, P. (Eds.). (2014). *Task sequencing and instructed second language learning*. Bloomsbury Publishing.

Cho, H. (2015). Effects of Task Complexity on English Argumentative Writing. *English Teaching*, *70*(2), 212.

Crookes, G., & Gass, S. M. (1993). *Tasks and Language Learning: Integrating Theory and Practice. Multilingual Matters 93*.

D'Argembeau, A., & Van der Linden, M. (2004). Phenomenal characteristics associated with projecting oneself back into the past and forward into the future: Influence of valence and temporal distance. *Consciousness and Cognition*, 13, 844–858.

D'Argembeau, A., Renaud, O., & Van der Linden, M. (2011). Frequency, characteristics and functions of future-oriented thoughts in daily life. *Applied Cognitive Psychology*, 25(1), 96-103.

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

Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer & S. Kornblum (Eds.), Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience (pp. 265-291). Cambridge, MA: MIT.

Guiraud, P. (1960). *Problèmes et méthodes de la statistique linguistique*. Presses universitaires de France.

Hunt, K. W. (1965). Grammatical Structures Written at Three Grade Levels. NCTE Research Report No. 3.

**CITED LITERATURE (Continued)**

Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing, 4*(1), 51-69.

Ishikawa, T. (2006). The effects of task complexity and language proficiency on task-based language performance. *The Journal of Asia TEFL*, 3(4), 193-225.

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical Sophistication as a Multidimensional Phenomenon: Relations to Second Language Lexical Proficiency, Development, and Writing Quality. *The Modern Language Journal*, *102*(1), 120-141.

Klinger, E. (1971). *Structure and functions of fantasy*. New York: Wiley.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing*, 17, 48-60.

Larsen-Freeman, D. (1978). An ESL index of development. *Tesol Quarterly*, 439-448.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, *16(3)*, *307-322.*

Long, M. H. (1985). *Bibliography of research on second language classroom processes and classroom second language acquisition* (No. 2). Center for Second Language Classroom Research, Social Science Research Institute, University of Hawaii at Manoa.

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

**CITED LITERATURE (Continued)**

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *Tesol Quarterly, 45(1), 36-62.*

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal, 96*(2), 190-208.

Lu, X. (2017). Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment. *Language Testing, 34*(4), 493-511.

Bygate, M., Skehan, P., & Swain, M. (2013). *Researching pedagogic tasks: Second language learning, teaching, and testing*. Routledge.

Mohammadzadeh Mohammadabadi, A. R., Dabaghi, A., & Tavakoli, M. (2013). The effects of simultaneous use of pre-planning along ± here-and-now dimension on fluency, complexity, and accuracy of Iranian EFL learners' written performance. *International Journal of Research Studies in Language Learning, 2*(3), 49–65.

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 University Press.

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.

Polio, C., & Shea, M. C. (2014). An investigation into current measures of linguistic accuracy in second language writing research. *Journal of Second Language Writing, 26*, 10-27.

Prabhu, N. S. (1987). *Second language pedagogy* (Vol. 20). Oxford: Oxford University Press.

**CITED LITERATURE (Continued)**

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

Richards, J., Platt, J., Weber, H., Inman, P., & Inman, P. (1986). Longman dictionary of applied linguistics. *RELC Journal,* 17(2), 105-110.

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, UK: Cambridge University.

Robinson, P. (2003a). Attention and memory in SLA. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 631-678). Oxford, UK: Blackwell.

Robinson, P. (2003b). The cognitive hypothesis of adult, task-based language learning. *Second Language Studies,* 21, 45-107.

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognitive hypothesis and second language learning and performance. *International Review of Applied Linguistics in Language Teaching, 45*(3), 161-176.

Ruiz-Funes, M. (2015). Exploring the potential of second/foreign language writing for language learning: The effects of task factors and learner variables. *Journal of Second Language Writing, 28,* 1-19.

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

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Basingstoke: Palgrave Macmillan.

**CITED LITERATURE (Continued)**

Sasayama, S. (2011). Cognition hypothesis and second language performance: Comparison of written and oral task performance. *University of Hawai'i Second Language Studies Paper 29 (2).*

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. *Perspectives on individual characteristics and foreign language education*, 6, 27.

Shajeri, E., & Izadpanah, S. (2016). The impact of task complexity along single task dimension on Iranian EFL learners' writing production. *Theory and Practice in Language Studies, 6, 935-945.*

Shehadeh, A., & Coombe, C. A. (Eds.). (2012). *Task-based language teaching in foreign language contexts: Research and implementation* (Vol. 4). John Benjamins Publishing.

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

Skehan, P. (Ed.). (2014). *Processing perspectives on task performance* (Vol. 5). John Benjamins Publishing Company.

Skehan, P., & Foster, P. (2001). Cognition and tasks. *Cognition and second language instruction,* 183-205.

Spreng, R.N., & Levine, B. (2006). The temporal distribution of past and future autobiographical events across the lifespan. *Memory and Cognition, 34*, 1644–1651.

Thomas, M., & Reinders, H. (Eds.). (2010). *Task-based language learning and teaching with technology*. A&C Black.

**CITED LITERATURE (Continued)**

Ure, J. (1971). Lexical density and register differentiation. *Applications of linguistics*, 443-452.

Van den Branden, K. (Ed.). (2006). *Task-based language education: From theory to practice*. Ernst Klett Sprachen.

VanPatten, B. (1990). Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition, 12*, 287-301.

Vyatkina, N. (2012). The development of second language writing complexity in groups and individuals: A longitudinal learner corpus study. *The Modern Language Journal, 96*(4), 576-598.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science, 3*(2), 159-177.

Willis, D. (1990). *The lexical syllabus: an new approach to language teaching*. Collins ELT.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, & complexity* (No. 17). University of Hawaii Press.

# APPENDICES

# Appendix A

LOW: (± causal reasoning)

Simple task: (- causal reasoning)

Write about your daily routine. Write about:

- What you eat and drink everyday,
- What time you do things (get up, got to university, go to bed, etc.), and
- What you do if you have free time

Complex task: (+ causal reasoning)

Write about an important person in your life. Say:

- Who the person is and where the person is from
- What the person does and where the person works and studies, and
- Why the person is important to you and why you like him/her.

INTERMEDIATE: (± here-and-now)

Simple task (+ here-and-now)

Write about a book or a TV program you like. Explain:

- What's it's about,
- Who the people are in it,
- When you read the book or when you watch the program, and
- What you like about it and why it is interesting.

Complex task (- here-and-now)

Write about what you want to do after you finish university. Write about:

- How you will celebrate finishing university
- Where you would like to live after you finish university
- What you will do for yourself and other people, and
- whether you will travel and why/why not.

# VITA

NAME: Abdullah Alshehri

EDUCATION: B.A., Languages and Translation, King Saud University, 2014.

ACADEMIC EXPERIENCE: Participation in the translation of *Profit Over People: Neoliberalism & Global Order* by Noam Chomsky

PROFESSIONAL MEMBERSHIPS: Illinois TESOL-BE
TESOL International Association