Unraveling the Effects of Task Sequencing on the Syntactic Complexity, Accuracy, Lexical Complexity, and Fluency of L2 Written Production

Mahmoud Abdi Tabari
University of Virginia

Michol Miller
Oklahoma State University

Abstract

Although several studies have explored the effects of task sequencing on second language (L2) production, there is no established set of criteria to sequence tasks for learners in L2 writing classrooms. This study examined the effect of simple−complex task sequencing manipulated along both resource-directing (± number of elements) and resource-dispersing (± planning time) factors on L2 writing compared to individual task performance using Robinson's (2010) SSARC model of task sequencing. Upper-intermediate L2 learners (*N = 90*) were randomly divided into two groups: (1) participants who performed three writing tasks in a simple–complex sequence, and (2) participants who performed either the simple, less complex, or complex task. Findings revealed that simple-complex task sequencing led to increases in syntactic complexity, accuracy, lexical complexity, and fluency, as compared to individual task performance. Results are discussed in light of the SSARC model, and theoretical and pedagogical implications are provided.

Résumé

Plusieurs études ont exploré les effets du séquençage des tâches sur la production en langue seconde (L2). Néanmoins, il n'y a pas eu d'ensemble de critères bien établis pour séquencer les tâches pour les apprenants dans les classes d'écriture de L2. Cette étude a examiné l'effet d'un séquençage simple−complexe manipulé de tâches ainsi que les facteurs de direction des ressources (± nombre d'éléments) et de dispersion des ressources (± temps de planification) sur l'écriture en L2 par rapport à la performance des tâches individuelles, et ce à l'aide du modèle de séquençage de tâches SSARC de Robinson (2010). Les apprenants de L2 de niveau intermédiaire-supérieur (*N = 90*) ont été répartis au hasard en deux groupes : (1) les participants qui ont effectué trois tâches d'écriture dans une séquence simple-complexe, et (2) les participants qui ont exécuté une des trois tâches : soit la simple, soit la moins complexe, soit la complexe. Les résultats ont révélé que le séquençage simple-complexe de tâches entraine une augmentation de la complexité syntaxique, de la précision, de la complexité lexicale et de la fluidité, par rapport à la performance des tâches individuelles. Les résultats sont discutés à la lumière du modèle SSARC et les implications théoriques et pédagogiques sont fournies.

Unraveling the Effects of Task Sequencing on the Syntactic Complexity, Accuracy, Lexical Complexity, and Fluency of L2 Written Production

Numerous studies in the task-based language teaching (TBLT) field have shown a growing interest in the constructs of cognitive task complexity (i.e., the inherent cognitive demands of a given task) and task sequencing over the past decade (Baralt, 2014; Lambert & Robinson, 2014; Levkina & Gilabert, 2014; Malicka, 2014, 2018). These TBLT studies relying on cognitive frameworks have examined theoretical predictions that sequencing tasks according to increasing complexity can maximize learning opportunities and provide optimal learning conditions. Building on Robinson's (2007) Triadic Componential Framework, recently renamed the SSARC (Stabilize, Simplify, Automatize, Restructure, Complexify) model of task sequencing (Robinson, 2010, 2015), most of these studies have produced relatively inconsistent findings regarding the effects of task complexity and task sequencing on second language (L2) oral production in the long and short term (Jackson & Suethanapornkul, 2013; Malicka & Sasayama, 2017). In the L2 writing context, this issue becomes more complicated because relatively few studies have investigated the effects of task sequencing on L2 writing, yielding mixed findings (Allaw & McDonough, 2019; Lambert & Robinson, 2014, Levkina & Gilabert, 2014). Thus, obscurities still exist on how to grade and sequence pedagogic tasks in L2 writing classrooms, highlighting the need for further investigation along this line of TBLT research. Furthermore, only two studies have so far examined the effect of task sequencing on L2 oral (Lambert & Robinson, 2014) and written (Allaw & McDonough, 2019) production by manipulating both resource-directing and resource-dispersing factors, yielding different results. To provide a more accurate and complete picture of the simultaneous interaction between resource-directing and resource-dispersing variables, further inquiries are warranted. Therefore, this study aims to investigate the effect of simple‒complex task sequencing manipulated along resource-directing (± number of task elements) and resource-dispersing (± planning time) factors on L2 written production when compared to individual task performance.

The theoretical framework of this study was based on Robinson's (2010, 2015) SSARC model of task sequencing which makes two principal claims: 1) decisions on task sequencing should be made according to cognitive factors, and 2) task sequencing should be performed in the simple-to-complex order. Although the SSARC model is primarily used for L2 oral production (Manchón, 2014; Tavakoli, 2014), Robinson has never limited its scope to speaking versus writing modalities, stating that the SSARC model is intended to apply across modalities when sequencing progressions of increasingly complex pedagogic versions of real-world target tasks. Only two researchers (Abdi Tabari, 2020; Allaw & McDonough, 2019) have tested the applicability of the SSARC model in predicting the effects of increasing task complexity on L2 writing and found that simple-complex sequencing of certain task types can be beneficial for L2 written production and development. Therefore, this study utilized Robinson's SSARC model as its theoretical basis and tested its predictive power in L2 writing. In order to situate this study in the current literature, a review of the Triadic Componential Framework, the SSARC model, and relevant TBLT research on the effect of task sequencing on L2 production is provided in the following section.

Background

Task Complexity and Task Sequencing: Framing Theoretical Perspectives

Over recent decades, many task complexity studies have been driven by two robust and competing models: Skehan's (1998, 2001) Limited Attentional Capacity (LAC) Model and Robinson's (2001, 2003) Cognition Hypothesis which make different predictions about the cognitive operations and attentional resources affecting L2 development. The LAC Model, grounded in psycholinguistic theories of first language (L1) acquisition, conceptualizes the relationship between cognitive and attentional resources during L2 processing (Skehan, 1998). This model predicts that increasing task complexity reduces cognitive capacity for monitoring linguistic form because complexity, accuracy, and fluency (CAF) compete intensely for the same attentional resources. Therefore, learners allocate attentional resources to CAF such that one of these constructs (e.g., accuracy) benefits at the expense of others (e.g., complexity and fluency), resulting in a tradeoff effect due to reduced cognitive capacity for monitoring formal aspects of the task (Abrams, 2019). The LAC model proposes that tasks should balance and distribute learner attention so that no CAF elements are neglected during L2 development. Conversely, Robinson's (2001, 2003) Cognition Hypothesis (CH), grounded in functional/cognitive linguistics, takes an alternate view of learners' cognitive abilities, arguing that learners possess multiple, rather than limited, attentional resources that do not compete. This forms the basis for the central claim of the CH which states that, since different CAF elements belong to different attentional resources, complexity and accuracy can be attended to concurrently with possible decays in fluency.

To provide a more comprehensive classification for determining task complexity, Robinson (2007) expanded on the CH with the Triadic Componential Framework (TCF), recently renamed the SSARC Model (Robinson, 2010). The TCF classifies task characteristics into three categories: task conditions, task difficulty, and task complexity. Task conditions represent interactional factors influencing the type and quantity of interactions required in a task, while task difficulty describes individual abilities and affective factors learners bring to task performance. Task complexity refers to task features that can be manipulated to challenge learners' cognitive resources and is further divided into resource-dispersing and resource-directing factors (see Table 1). Resource-directing factors direct learner attention to language needed for task completion, while resource-dispersing variables make procedural and performative cognitive demands. Robinson (2001, 2003) stipulates that simple tasks yield greater fluency as resource-dispersing variables become more complex, while further complexification along resource-directing variables yields greater accuracy. Such complexification increases linguistic production and pushes learners to adjust and expand their interlanguage but decreases fluency.

Table 1

Task Variables in Robinson's Triadic Componential Framework (Adapted from Robinson & Gilabert, 2007, p. 164)

Task Complexity (Cognitive Factors)	Task Conditions (Interactive Factors)	Task Difficulty (Learner Factors)
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	+/- open solution	h/l working memory
+/- few elements	+/- one-way flow	h/l reasoning
-/+ spatial reasoning	+/- convergent solution	h/l task-switching
-/+ causal reasoning	+/- few participants	h/l aptitude
-/+ intentional reasoning	+/- few contributions needed	h/l field independence
/+ perspective-taking	+/- negotiation not needed	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	+/- same proficiency	h/l openness to experience
+/- single task	+/- same gender	h/l control of emotion
+/- task structure	+/- familiar	h/l task motivation
+/- few steps	+/- shared content knowledge	h/l processing anxiety
+/- independency of steps	+/- equal status and role	h/l willingness to communicate
+/- prior knowledge	+/- shared cultural knowledge	h/l self-efficacy

Several studies have investigated the effects of task complexity on learners' CAF by manipulating resource-directing and resource-dispersing factors. Relevant to the present research are studies examining the effect of raising task complexity along the resource-directing variable of number of elements (± elements) and the resource-dispersing variable of pre-task planning time (± planning) on L2 written production.

Studies investigating raised task complexity along ± elements in L2 writing have produced mixed findings. Some reported that raising task elements promotes accuracy (Kuiken & Vedder, 2011, 2012) or lexical diversity (Lee, 2018) or both accuracy and

lexical diversity (Kuiken & Vedder, 2007, 2008). Conversely, Lee (2018) reported that greater task complexity along ± elements decreased accuracy and syntactic complexity. Studies examining increased task complexity by reducing or excluding planning time have also produced mixed results. Four studies reported both decreased accuracy and syntactic complexity (Abdi Tabari, 2020; Ellis & Yuan, 2004; Farahani & Meraji, 2011; Rostamian et al., 2018), while Rahimi (2018) reported increased syntactic complexity. Regarding lexical diversity, both increases (Ong & Zhang, 2010; Rostamian et al., 2018) and decreases (Abdi Tabari, 2020) have been found. Finally, greater task complexity along ± planning induced decreased fluency in four studies (Ellis & Yuan; Farahani & Meraji, 2011; Rahimi, 2018; Rostamian et al., 2018), but increased fluency in two studies (Abdi Tabari, 2020; Ong & Zhang, 2010). Other studies examining increased task complexity in L2 writing manipulated both ± elements and ± planning alongside other variables. Raising task complexity along ± elements and ± reasoning produced mixed effects: Rahimi (2018, 2019) reported decreased accuracy but increased syntactic complexity, while Frear and Bitchener (2015) reported decreased syntactic complexity but increased lexical diversity. Rahimi (2018) found that raising complexity along ± planning, ± elements and ± reasoning yielded increased syntactic and lexical complexity but reduced fluency in written tasks.

Expanding on earlier theories of task complexity, Robinson's (2010) SSARC model distinguishes which task variables should be manipulated during task sequencing, favouring cognitive factors over interactional and learner factors. This comprises the first principle in the SSARC model: Tasks should be sequenced only according to cognitive complexity, operationalized as resource-directing and resource-dispersing variables, while other variables remain constant. The second task sequencing principle states that task complexity should first be increased along resource-dispersing variables, followed by resource-directing variables (Robinson, 2010).

The SSARC model illustrates the rationale for these two principles. In Step 1, Stabilize and Simplify, learners complete simple tasks, engaging their current interlanguage. In Step 2, Automatize, complexity is increased along resource-dispersing variables to encourage quicker access to learners' interlanguage. In Step 3, Restructure and Complexify, complexity is raised along both resource-dispersing and resource-directing variables so that learners' interlanguage systems are destabilized and restructured, thus promoting more complex interlanguage (Robinson, 2010). Therefore, the SSARC model offers clear pedagogical implications for task-based syllabus design by proposing a pre-determined sequence that promotes interlanguage development, as shifts in task complexity induce gradual shifts in interlanguage throughout the task sequence.

Review of Empirical Studies: Effects of Task Sequencing on L2 Production

To test Robinson's SSARC Model, several studies investigating the effect of task complexity and sequencing on L2 development have been conducted on L2 oral and written production by raising task complexity along resource-directing and/or resource-dispersing variables over different sequencing orders.

Studies testing the SSARC model by manipulating only resource-directing variables have reported a range of findings. Levkina and Gilabert (2014) tested the effect of different sequences (simple-complex, complex-simple, and randomized) complexified along two variables (± spatial reasoning; ± perspective-taking) on learners' retention of

spatial expressions over time. Immediate post-tests revealed that complex-to-simple sequencing led to greater short-term retention of spatial expressions, while delayed post-tests showed that simple-complex sequencing resulted in greater long-term retention. Baralt (2014) examined the effects of four simple-complex sequences (SSC, SCS, CSC, CCS) on L2 oral and written production by raising complexity along ± reasoning demands. Results showed that sequences with more complex tasks (CCS and CSC) generated increased learning opportunities and greater L2 development. Malicka (2014, 2018) tested the effects of different sequences (simple-to-complex, randomized, or individual tasks) on L2 oral production by raising complexity along ± reasoning demands and ± few elements. Both of her studies reported that simple-to-complex task sequencing led to improved speech rate, accuracy, and structural complexity. While the studies above indicated that simple-to-complex sequencing is effective, the SSARC model was not fully tested since only resource-directing variables were manipulated.

Only two studies have examined the effect of task sequencing by manipulating both resource-directing and resource-dispersing factors according to the SSARC model's proposed order. Lambert and Robinson (2014) explored the effects of simple-complex and randomized task sequencing on L2 written production, modifying both resource-directing (± few elements; ± reasoning demands) and resource-dispersing variables (± planning; ± prior knowledge; ± number of steps; ± multi-tasking). Findings indicated the simple-to-complex task sequence led to greater overall long-term benefits. More recently, Allaw and McDonough (2019) tested the effect of simple-complex versus complex-simple sequences on L2 written production by manipulating both resource-directing (± spatial reasoning) and resource-dispersing variables (± task structure). While both sequences yielded increased lexical diversity, grammatical accuracy, and fluency, the simple-complex sequence led to greater overall performance and long-term improvement. This provides strong support for Robinson's SSARC model, as the model's proposed simple-complex sequence yielded the greatest performance gains.

The Present Study

The current study was motivated by two gaps identified in the TBLT literature: First, several studies have explored the effects of task sequencing on L2 oral production (Malicka, 2014, 2018), acquisition (Baralt, 2014), and interaction (Kim & Payant, 2014), but the limited number of studies investigating the impacts of task sequencing on L2 writing yielded contradictory results (Allaw & McDonough, 2019; Lambert & Robinson, 2014, Levkina & Gilabert, 2014). Second, while some TBLT studies have explored task sequencing by manipulating either resource-directing or resource-dispersing factors, only two task sequencing studies have simultaneously manipulated both of these task complexity factors as proposed in the SSARC model (Allaw & McDonough, 2019; Lambert & Robinson, 2014), producing different results. Thus, further exploration is warranted to provide additional empirical evidence for the SSARC model. To address the aforementioned gaps, the current study seeks to answer the following question: What is the effect of simple–complex task sequencing on the syntactic complexity, accuracy, lexical complexity, and fluency of L2 written production when compared to individual task performance?

Methodology

Participants and Context

This study included a sample of 90 undergraduate learners (46 women and 44 men) who enrolled in a fifteen-week, advanced multilingual writing course at a large university in the United States. They were recruited from five different classes taught by two different instructors using a communicative-based syllabus (Ellis, 2003) and employing similar teaching materials. Participants were aged between 19-21 years old ($M = 20.6, SD = 8.61$) and had lived in English-speaking countries for three years ($M = 2.6, SD = .8$). The participants came from a variety of L1 backgrounds. Half of the participants were Hispanic ($N = 45$) while the rest were from China ($N = 8$), South Korea ($N = 7$), France ($N = 6$), Japan ($N = 5$), Scandinavia ($N = 5$), Spain ($N = 4$), Russia ($N = 3$), Iran ($N = 2$), Poland ($N = 2$), Holland ($N = 2$), and Nigeria ($N = 1$). They had learned English as an L2 for 9 to 12 years ($M = 10.3, SD = 4.28$) at the time of data collection. Participants received nearly 39 hours of formal classroom teaching before data collection and all had upper-intermediate proficiency levels, based on their TOEFL-iBT scores in the 17-23 range and performance on a writing placement test administered by the university annually. After signing informed consent forms, participants were randomly divided into two groups to either perform written tasks in a simple−complex sequence or complete the same tasks individually.

Procedure

Prior to starting the main experiment, a pilot study was conducted to determine the time needed for performing different versions of a written decision-making task and validate the assumptions about task complexity manipulations before the main experiment. Thus, 30 participants first performed a simple decision-making task with 10 minutes of planning time, then the same task without planning time, and finally a complex version of the task without planning time. After performing each task, the participants completed the task complexity questionnaire to validate the task complexity manipulations. The time each participant spent on different versions of the task was recorded by the researcher, for which descriptive statistics are presented in Table 2.

Table 2

Means and Standard Deviations for the Length of Time Spent on Each Task Performance

| | Task sequencing | | | | | |
| | Simple + Planning | | Simple − Planning | | Complex − Planning | |
	M	SD	M	SD	M	SD
Length of time for task performance	17.2	2.8	23.8	3.11	30.2	3.76

As shown in Table 2, there was a gradual pattern of increasing lengths of time spent on the written tasks as cognitive complexity increased, such that the simple task with planning time was the shortest task to perform, followed by the simple task without planning, and the complex task without planning time as the longest. To ensure that pre-task planning was properly operationalized, the average amount of time spent by participants to perform the tasks was used as the time limit in the main writing experiment.

After conducting the pilot study, the researcher visited regularly scheduled multilingual writing classes and randomly assigned participants to two groups to conduct the main experiment: 1. the simple–complex sequencing group (*n = 30*), and 2. the individual task group (*n = 60*). Following the SSARC model of task sequencing and the times set in the pilot study, the first group performed the written tasks in the simple-to-complex order with 5-minute intervals; participants performed the simple task in 17.2 minutes with 10 minutes of planning time, then the same task in 23.8 minutes without planning time, and finally the complex one in 30.2 minutes without planning time. Similar to Malicka's (2018) study, 60 participants in the individual task group were subdivided into three groups of equal size (*n = 20*); each subgroup completed only one task under the same task conditions as in the sequencing group. Thus, the difference between the two groups is that in the simple– complex group, all participants performed the tasks successively, while in the individual task group, participants in each subgroup completed only one task at a pre-determined cognitive complexity level.

Writing Tasks

In line with the principles of the SSARC model, different versions of a decision-making task varying in terms of inherent cognitive complexity were created and sequenced in the simple–complex order. The versions of the writing task, manipulated along resource-directing and resource-dispersing factors, were different in terms of cognitive complexity by decreasing or increasing the number of task elements (a resource-directing factor) and providing or removing pre-task planning time (a resource-dispersing factor). The decision-making task included specific descriptions of different job candidates who applied for a software engineering position at a well-known company; the number of candidates varied in different versions of the written task. Table 3 summarizes the sequenced writing task in light of the SSARC model.

Table 3
The Sequence of Different Versions of the Writing Task in Accordance with the SSARC Model

Task complexity factors	Task sequencing		
	Stage (I): Simple	Stage (II): Less complex	Stage (III): Complex
Resource-directing: the number of elements	− Elements	− Elements	+ Elements
Resource-dispersing: planning time	+ Planning time	− Planning time	− Planning time

In Stage I, participants were given four job candidates' application dossiers, and based on the information given, had to decide which two candidates would be the most qualified for the company's software engineering position. Before writing, they had 10 minutes of planning time to inspect each candidate's application dossiers and write some planning notes, allowing participants to prepare for the task performance and focus on content, language, and organization. More importantly, the provision of planning time was hypothesized to mitigate the online processing load during writing and free up attentional resources for focusing on different aspects of the task; furthermore, the task consisted of

only four elements placing a lower demand on their working memory. The combination of a simple version of the task and planning time was intended to simplify input corresponding to the first stage in the SSARC model.

In Stage II, the number of task elements remained unchanged and participants again decided which two out of four candidates would be top-tier candidates for the software engineering position. Following the second stage in the SSARC model, in which tasks should be cognitively demanding on resource-dispersing factors (Robinson, 2010), planning time was removed to increase the cognitive complexity on this version of the task, in order to improve automatization of the writing process. Thus, participants had to produce similar ideas as in the simple task, while engaging in planning and writing at the same time. Compared to the simpler task, this task was considered more cognitively demanding since participants wrote without planning time.

In Stage III, participants were given six job candidates' application dossiers and had to decide which two would be the most qualified candidates for the company's open position. The increase in the number of elements along with the removal of planning time bolstered the complexity of the writing task such that participants had to simultaneously process and analyze six candidates, then plan and organize their arguments before finally expressing their new ideas. This increase in cognitive complexity corresponds to stage three in the SSARC model postulating that tasks should be cognitively demanding along both resource-directing and resource-dispersing factors, connecting learners to novel linguistic forms, and pushing them to complexify their interlanguage (Robinson, 2010).

Validation of Cognitive Task Complexity Assumptions

TBLT researchers have called for examining the validity of task complexity assumptions to confirm that tasks designed to be cognitively complex actually result in varying levels of cognitive complexity for learners (Norris, 2010; Révész, 2014). In response, numerous studies have used various subjective and objective techniques to verify the assumptions about the impacts of task manipulations on task complexity, as summarized in Table 4.

Table 4
Techniques for Validating Cognitive Task Complexity

Validation Technique	Studies
Learner self-ratings of task difficulty	Allaw & McDonough (2019); Frear & Bitchener (2015)
Expert judgments of task difficulty	Révész, Michel, & Gilabert (2015); Rahimi (2019)
Time estimations	Baralt (2013); Malicka & Levkina (2012)
Time-on-task	Lee (2018)
Eye-tracking	Michel & Smith (2018); Révész, Sachs, & Hama (2014)
Dual task methodology	Révész, Michel, & Gilabert (2015), Sasayama (2016)
Stimulated recall protocols	Michel, Révész, Lu, Kourtali, Lee, & Borges (2020); Malicka, 2018
Writing measures (fluency, pausing, and revision behaviours)	Révész, Kourtali, & Mazgutova (2017)

Before the main experiment, a pilot study with 30 students and 10 teachers was conducted to validate the speculated differences in cognitive complexity levels between the written tasks via two different analytical, subjective techniques: learner self-ratings and expert judgments. To examine learner perceptions and attitudes about the perceived difficulty, stress, and cognitive load of the tasks, a nine-point Likert-scale questionnaire adapted from Lee's (2018) original questionnaire was used to elicit participants' responses to 11 items representing different categories. In addition, all participants were requested to answer four open-ended questions: 1. Did you notice any difference between performance of the writing task with the provision of planning time versus no provision of planning time? 2. Which one was more difficult to perform and why? 3. In your opinion, was there any difference between the two writing tasks in terms of difficulty? 4. Which of the two versions of the written task was more difficult to perform and why? As presented in Table 5, the descriptive results demonstrated a steady increase in participants' self-ratings of perceived difficulty, mental effort, and stress as cognitive complexity intensified. In line with the predictions made, the overall ratings for the simple versions of the task were lower than those for the complex task. In addition, the simple task with planning time obtained lower means than the version without planning time in terms of the three affective variables. Regarding participants' responses to the follow-up questions, one student wrote,

> The preparation time before writing was really accommodating and beneficial because it helped me gather my thoughts create a clear outline and write with more comfort and confidence. But I had a hard time performing the second task without planning time and it became much harder while doing the third task including more job candidates. Compared to the third task, the second one was less difficult because I already carried it out and had some familiarity with the task performance.

Table 5
Descriptive Statistics for Participants' Self-ratings

Affective variables	Task sequencing					
	Simple + Planning		Simple − Planning		Complex	
	Mean	SD	Mean	SD	Mean	SD
Difficulty	3.83	1.41	4.12	1.55	4.63	1.13
Mental Effort	3.45	1.15	4.39	1.03	5.10	1.72
Stress	3.18	1.46	3.69	1.15	4.22	1.63

To see how experts judge the cognitive load of tasks, 10 university instructors with considerable experience teaching L2 writing courses were asked to rate the difficulty of tasks on a Likert scale originally adapted from Robinson's (2001) questionnaire. They were also provided with open-ended questions regarding the overall perceived difficulty and mental effort required to complete the tasks. Their responses to the questionnaire items were analyzed by one rater and the open-ended responses were examined by two raters. The results are presented in Table 6. As expected, students' and instructors' ratings were consistent: the complex task was rated higher than the simple task, and the simple task with no planning time was rated higher than the simple task with planning time.

Table 6
Descriptive Statistics for Experts' Judgment

Affective variables	Task sequencing					
	Simple + Planning		Simple − Planning		Complex	
	Mean	SD	Mean	SD	Mean	SD
Difficulty	3.48	1.35	3.94	1.15	4.69	1.41
Mental Effort	3.79	1.22	4.10	1.48	4.86	1.27

One expert wrote,

> In my opinion, the final task was the difficult one because it would require learners to engage in processing and analyzing in more depth. In addition, they would have to choose two out of six candidates with strong applications and support their choices with evidence. However, I would perceive the first task aligned with planning time as the least difficult task because learners can prepare their written plans beforehand and use them while writing. The same version of the task without planning is the medium task which inevitably requires more mental effort and time to process and perform due to the absence of planning time.

CALF Measures

Following Norris and Ortega (2009), who argue for assessing syntactic complexity multidimensionally while warning against using redundant measures which result in a multicollinearity effect, four different measures were utilized to gauge different subconstructs of complexity: 1. the mean length of T-unit (MLT) as a general measure of syntactic complexity calculated by dividing the total number of words by the total number of T-units in a text, 2. the ratio of dependent clauses to T-units (DC/T) as a measure of subordination complexity gauged by the ratio of dependent clauses to T-units in a text, 3. the number of complex nominals per T-unit (CN/T) analyzed by dividing the total number of complex nominals by the total number of clauses, and 4. the number of complex nominals per clause (CN/C) computed by the total number of complex nominals by the total number of T-units (Lu, 2010). The last two measures (CN/T and CN/C) were chosen to examine phrasal complexity in participants' written production. In our analysis, T-unit was used rather than C-unit or AS-unit because the nature of written tasks was monologic (Foster, Tonkyn, & Wigglesworth, 2000). It is defined as "one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it" (Hunt, 1970, p. 4).

In line with Ellis and Yuan (2004), the accuracy of L2 written production was examined by calculating the ratio of error-free clauses to the total number of clauses and the ratio of correct verb forms to the total number of verbs used in each text. To calculate error-free clauses, participants' written production was divided into clauses and lexical, morphological, and syntactic errors were identified and marked. Any unmarked clause was considered error-free. For each participant, the proportion of error-free clauses was regarded as the resulting score. Given that the error-free clauses metric is a holistic measure of accuracy, correct verb forms were used as a specific measure. For each

participant, the proportion of verbs free of tense, aspect, modality, or agreement errors was used as a score of analysis.

Lexical complexity was gauged with respect to diversity and sophistication. The measure of textual lexical diversity (MLTD) was used by computing "the mean length of sequential word strings in a text that maintain a given TTR value" (Mazgutova & Kormos, 2015, p. 5). MLTD was used rather than other vocabulary diversity metrics, specifically mean segmental type-token ratio (MSTTR), because it is least affected by text length (Mazgutova & Kormos, 2015; McCarthy & Jarvis, 2010) and therefore is considered a more reliable indicator of vocabulary. Additionally, there is a growing demand for using lexical sophistication metrics in measuring L2 written production (Johnson, 2017). In response, the log frequency of content words was utilized to examine the lexical sophistication of texts and was calculated by the ratio of the log frequency to content words in the CELEX database (McNamara et al., 2014). There were two main reasons for the selection of the log frequency of content words: 1. compared to frequency band measures, the log frequency of content words can better represent large and small improvements in the participants' written production due to inclusion of the frequency counts from a large corpus (Kyle & Crossley, 2015); and 2. this metric measures lexical sophistication with a higher degree of reliability in comparison to the raw frequency of content words (Kormos, 2011). Finally, fluency was calculated by counting the number of words produced within a set time (Abrams, 2019). This product-based measure was selected for two reasons: 1. it has ecological validity such that teachers can use it in curriculum-based assessment, and 2. it allows comparability of the results with the findings of past studies.

Statistical Analysis

In the analyses of written performances, we used nine different measures to examine the effects of task sequencing as an independent variable on different constructs of L2 written production as dependent variables. First, means and standard deviations were calculated for all the variables in the different groups. Then, data sets were directly imported into R version 3.6.1 (R Development Core Team, 2019) to check the normality assumption through normal Q–Q plots and the Kolmogorov–Smirnov test. Given that six out of nine variables were associated with a violation of the normality assumption, non-parametric statistics (i.e., Mann-Whitney U test) were used to answer the research question. The Mann-Whitney U test was performed to detect significant comparisons between the simple‒complex sequencing group and the individual task group. The level of significance for this study was set at an alpha level of 0.05. For the Mann-Whitney U test, Cohen's d was employed to measure effect sizes. Following Plonsky and Oswald's (2014) benchmarks, d values of .40, .70 and 1.00 were considered as small, medium, and large, respectively.

Results

Effects on Syntactic Complexity

The descriptive statistics for the measures of general complexity (MLT), subordination complexity (DC/T), and phrasal complexity (CN/T and CN/C) are presented for the simple–complex sequencing group and the individual task group in Figure 1. The inferential statistics revealed significant differences in performance between the two groups with respect to MLT in the simple task with planning time, MLT and DC/T in the simple task without planning time, and DC/T in the complex task. The participants in the simple–complex sequencing group produced more complex structures than their counterparts in the individual task group on two simple tasks with and without planning (simple task + planning, $p = .000$ and simple task − planning, $p = .000$). In both cases, the d scores indicated large effect sizes corroborating a substantial difference between the two groups' performance in the case of MLT. Similar results were also found between the simple–complex sequencing group and the individual task group in terms of complex subordinations in the simple task without planning time ($p = .000$) and the complex task ($p = .000$) with large effect sizes ($d = 1.08$ and 1.03, respectively). Table 7 displays the performance of the simple–complex sequencing group as opposed to the individual task group for complexity subconstructs.

Table 7

Inferential Statistics for Different Subconstructs of Syntactic Complexity

Dependent variable	Group	Simple + Planning				Simple − Planning				Complex			
		P	95% CI		d	P	95% CI		d	P	95% CI		d
Complexity			Lower Bound	Upper Bound			Lower Bound	Upper Bound			Lower Bound	Upper Bound	
MLT	IND	.000**	0.68	2.43	1.00	.000**	0.75	2.99	1.08	.100	-0.20	1.65	.45
	SCS												
DC/T	IND	.160	-0.02	0.08	.41	.000**	0.04	0.15	1.08	.000**	0.04	0.13	1.03
	SCS												
CN/T	IND	.930	-0.19	0.12	.03	.400	-0.14	0.30	.24	.920	-0.11	0.26	.03
	SCS												
CN/C	IND	.340	-0.17	0.26	.27	.490	-0.12	0.29	.20	.600	-0.06	0.14	.15
	SCS												

Figure 1
Mean Scores for MLT, DC/T, CN/T, and CN/C
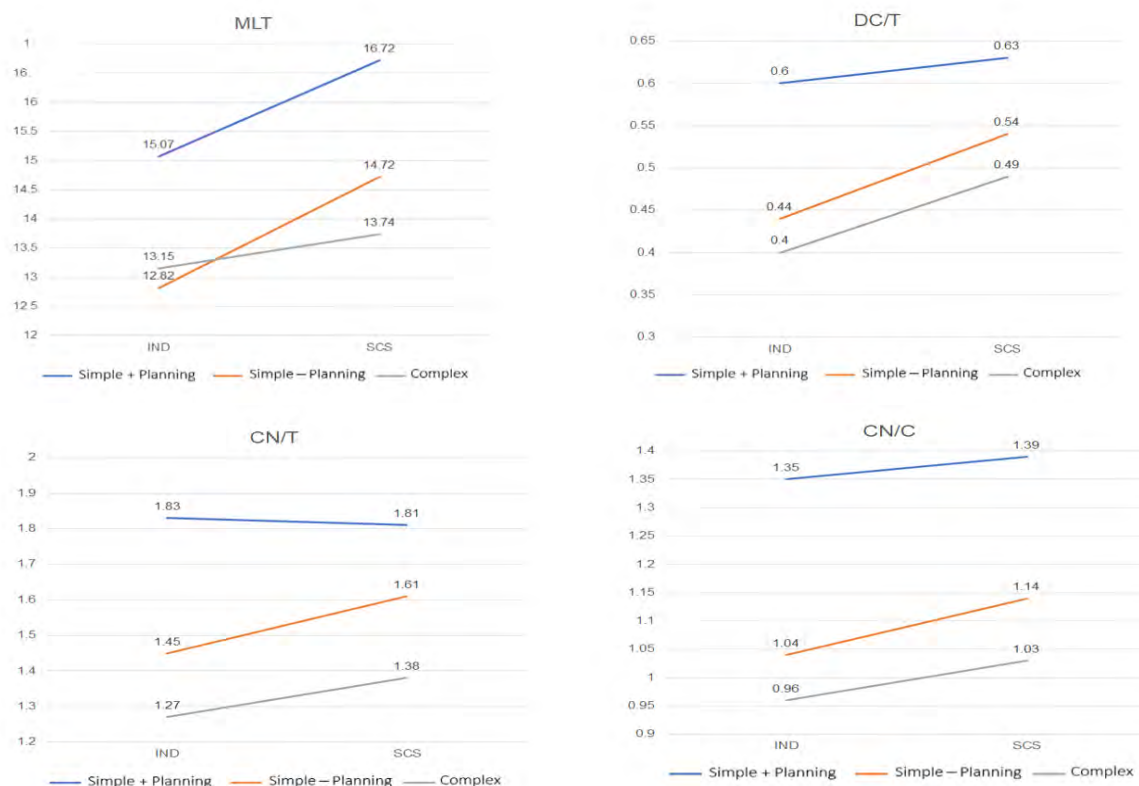
Note. IND = individual task group, SCS = simple–complex sequencing group
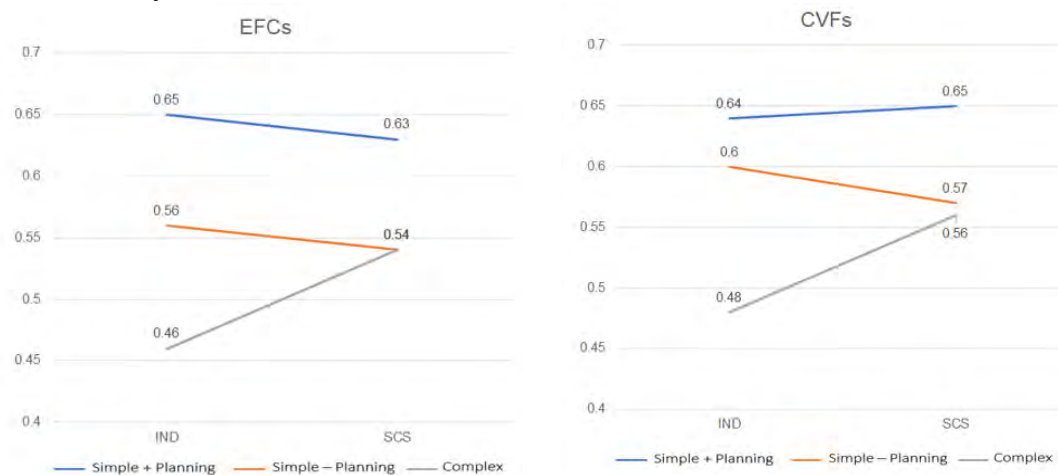
Effects on Accuracy

        The descriptive statistics revealed that the simple–complex sequencing group yielded higher mean scores than the individual task group regarding the two accuracy measures in the complex task, but the opposite was found for the simple task without planning time. In the case of the simple task with planning time, the simple–complex sequencing group produced a higher mean score than the individual task group on correct verb forms; however, the latter had a higher mean score on error-free clauses (see Figure 2). The results from the Mann-Whitney *U* test revealed that, whereas the individual task group outscored the simple–complex sequencing group regarding error-free clauses and correct verb forms in the simple tasks with and without planning time, the comparisons did not reach statistical significance. The *d* scores showed small effect sizes between the two groups' performance in the simple task with planning time (*d* = .2 for EFCs and *d* = .17 for CVFs) and in the simple task without planning time (*d* = .21 for EFCs and *d* = .12 for CVFs). Nonetheless, the proportion of error-free clauses and correct verb forms in the simple–complex sequencing group was higher than in the individual task group in the complex task and the comparisons were statistically significant with large effect sizes (*p* = .000, *d* = .95 in the case of EFCs and *p* = .001, *d* = .86 in the case of CVFs). The inferential statistics are summarized in Table 8.

Table 8

Inferential Statistics for Different Subconstructs of Accuracy

Dependent variable	Group	Simple + Planning				Simple − Planning				Complex			
		P	95% CI		*D*	*P*	95% CI		*d*	*P*	95% CI		*d*
			Lower Bound	Upper Bound			Lower Bound	Upper Bound			Lower Bound	Upper Bound	
Accuracy													
EFCs	IND	.480	-0.10	0.06	.20	.470	-0.06	0.04	.21	.000**	0.03	0.12	.95
	SCS												
CVFs	IND	.570	-0.05	0.08	.17	.670	-0.06	0.03	.12	.001**	0.02	0.11	.86
	SCS												

Figure 2
Mean Scores for EFCs and CVFs

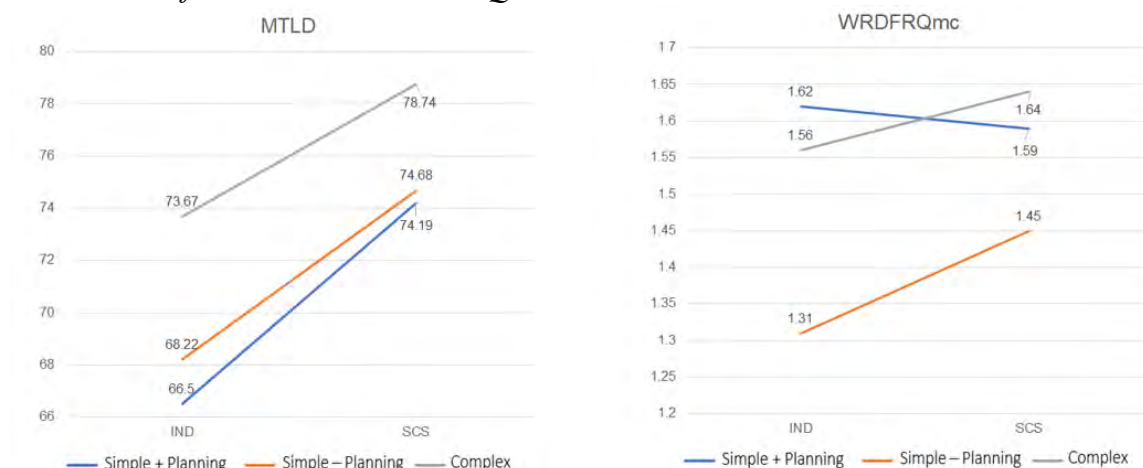*Note.* IND = individual task group, SCS = simple−complex sequencing group

Effects on Lexical Complexity

The descriptive statistics showed the simple−complex sequencing group produced higher mean scores than the individual task group with respect to lexical diversity (MTLD) and sophistication (WRDFRQmc) in the three writing tasks (see Figure 3). The comparisons between the two groups' performance reached statistical significance only in the case of MTLD with a medium effect size in the simple task (p = .004, d = .60). In addition, significant comparisons were found between the simple−complex sequencing group and the individual task group regarding MTLD and WRDFRQmc in the complex task. The d scores indicated a medium effect size for the simple−complex sequencing group compared to the individual task group (d = .66 for MTLD and d = .63 for WRDFRQmc). Table 9 presents the inferential statistics for two measures of lexical complexity.

Table 9

Inferential Statistics for Different Subconstructs of Lexical Complexity

Dependent variable	Group	Simple + Planning				Simple − Planning				Complex			
		P	95% CI		*D*	*P*	95% CI		*d*	*P*	95% CI		*d*
Lexical complexity			Lower Bound	Upper Bound			Lower Bound	Upper Bound			Lower Bound	Upper Bound	
MTLD	IND	.004*	0.37	13.56	.60	.120	-1.39	12.05	.45	.003*	0.70	9.47	.66
	SCS												
WRDFRQmc	IND	.610	-0.17	0.10	.15	.148	0.00	0.23	.57	.003*	0.01	0.29	.63
	SCS												

Figure 3
*Mean Scores for MTLD and WRDFRQmc*

Note. IND = individual task group, SCS = simple–complex sequencing group

Effects on Fluency

The descriptive statistics showed that participants in the simple–complex sequencing produced more words within the designated time limit in the simple task with planning time and in the complex task when compared to the individual task group. Nevertheless, the latter generated more words in the simple task without planning time than the former (see Figure 4). As can be observed in Table 10, the results revealed that the comparison between the two groups was statistically significant only in the case of the simple task with planning time ($p = .001$) and this was reflected by a large effect size ($d = .87$). However, no comparisons between the two groups were found to be statistically significant with respect to fluency in the simple task without planning time ($p = .350$) and in the complex task ($p = .320$).

Table 10

Inferential Statistics for Fluency

Dependent variable	Group	Task sequencing										

Dependent variable	Group	Simple + Planning				Simple − Planning				Complex			
		P	95% CI		*d*	*P*	95% CI		*d*	*P*	95% CI		*d*
Fluency			Lower Bound	Upper Bound			Lower Bound	Upper Bound			Lower Bound	Upper Bound	
W/T	IND	.001**	6.00	28.00	.87	.350	-31.00	17.00	.27	.320	-12.00	39.00	.29
	SCS												

Figure 4
Mean Scores for W/T
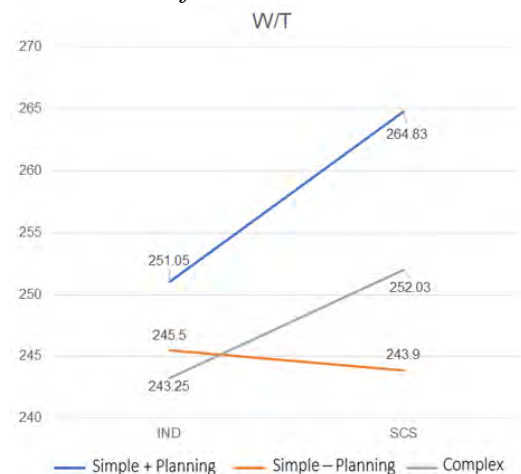
Note. IND = individual task group, SCS = simple–complex sequencing group

   Overall, these results revealed that the simple–complex sequencing group produced greater syntactic complexity at the general level (simple tasks with and without planning time) and at the subordination level (simple task without planning and complex task); wrote more error-free clauses and correct verb forms (complex task); exhibited more diverse and sophisticated vocabulary (simple task with planning and complex task); and finally, wrote faster and generated more words within a set time (simple task with planning time) as compared to the individual task group.

Discussion

   The major thrust of this study was to explore the effects of task sequencing on multilingual learners' written production in accordance with the SSARC model. Given that past studies tested the role of the SSARC model in L2 oral production (Baralt, 2014; Malicka, 2014, 2018), finding that the simple-to-complex order led to significant gains in L2 production over the short or long term, the current study examined the effectiveness of the SSARC model in L2 written production using a simple–complex task design. The results demonstrated that simple–complex task sequencing favoured syntactic and lexical complexity, promoted accuracy, and assisted fluency, providing empirical evidence supporting the theoretical claim of the SSARC model regarding the beneficial role of simple–complex task sequencing for L2 written production.
   With respect to syntactic complexity, whereas production of complex structures gradually decreased in both the simple–complex sequencing and individual performance groups over the sequence, the former produced both a significantly higher mean length of T-unit in the sequence's first two tasks, characterizing general complexity, and significantly more subordination, recognized as the most indicative source of complexification at the intermediate level, in the last two tasks in the sequence. Compared to the individual task group, the simple–complex sequencing group also produced more complex nominal structures throughout the sequence, manifesting greater phrasal-level complexification as a result of increased cognitive complexity, although not at a

statistically significant level. These results partially corroborate Malicka (2014, 2018) which also found that simple-complex sequencing increased syntactic complexity at the clausal level, but conversely reported decreases in subordination. Furthermore, the results revealed that in the simple task with planning time, only general syntactic complexity was significantly fostered in the simple–complex sequencing group. As reported previously (Abdi Tabari, 2020; Ellis & Yuan, 2004; Farahani & Meraji, 2011; Rostamian et al., 2018), pre-task planning provided learners with the opportunity to plan, linguistically encode their messages, and produce more complex writing. Increases in general complexity and complex subordinations can also be explained by the SSARC model, postulating that removing pre-task planning time and increasing cognitive complexity induced learners to exhibit higher levels of syntactic complexity because their exposure to the simple task created more scaffolded opportunities for rehearsal, serving as a preparatory mechanism to complexify their writing at a deeper level (Malicka, 2014). Nevertheless, increases in the number of elements did not prompt learners to stretch their syntactic resources in the complex task to produce significantly more complex structures at the phrasal complexity level, possibly due to their proficiency level. Overall, these results partially support the Cognition Hypothesis (Robinson, 2003) which predicts that increasing task complexity along resource-directing factors can push learners to extend their existing L2 repertoire to meet task demands.

Regarding L2 writing accuracy, the results revealed that on both measures for this construct, correctness of linguistic forms steadily decreased in both groups' performance, with the simple task with planning generating the most accurate forms. Nevertheless, significant differences were observed only in the complex task resulting in increased accuracy in the simple–complex group. These results suggest that when tasks are performed in the simple-to-complex order, learners can produce more correct clauses and verbs and improve the accuracy of their written production as compared to tasks performed in isolation. These results echo findings of past oral- and written-performance task sequencing studies (Allaw & McDonough, 2019; Malicka, 2014, 2018) reporting increases in grammatical accuracy under the simple–complex task condition. Following the SSARC model, the simple task with planning time may provide learners with the opportunity to direct more attention to linguistic forms, activate their monitoring behaviour, and rehearse target-like structures. In the simple task without planning time and the complex task, they can recall those structures, be cognizant of problematic linguistic areas, and avoid errors although pre-task planning time is removed. Consequently, less deviation from target-like structures should occur on the less complex and complex versions of the task. These results support Robinson's Cognition Hypothesis which postulates that complex tasks can lead to more accurate language if complexification occurs along resource-directing factors. However, it is imperative to mention that the results of this study only provide some preliminary evidence confirming the prediction of dual increases in complexity and accuracy.

Concerning lexical complexity, results showed that lexically diverse items gradually increased in both groups' performance and the complex task triggered the highest lexical diversity. Regarding lexical sophistication, a different pattern was found for the two groups such that the simple–complex sequencing group produced the most sophisticated vocabulary in the complex task, but the individual task group generated the most sophisticated lexical items in the simple task with planning. These results support

previous findings that increased task complexity induces greater lexical diversity in both oral and written performance (Abdi Tabari, 2020; Frear & Bitchener, 2015; Kuiken & Vedder, 2007, 2008; Levkina & Gilabert, 2012; Ong & Zhang, 2010; Rahimi, 2018), as well as Allaw and McDonough's (2019) findings, providing further confirmation of the SSARC model's claim that increasing cognitive complexity in the simple‒complex sequence induces greater lexical complexity. The simple task helped learners focus on simplified input and stabilize recently learned lexical items, while also creating prerequisite conditions for the subsequent complexification of their lexical production. The less complex task, removing pre-task planning, encouraged greater independence in using a wider range of lexical items, as well as fostering consolidation and automatization. Finally, the complex task, increasing the number of elements, aided the learners in complexifying their lexical production over the simple-to-complex sequencing order.

Regarding L2 writing fluency, both groups generated the highest number of words at the fastest speed in the simple task with planning time. Notably, under the individual task performance condition, the proportion of words produced and writing fluency decreased steadily as a result of increasing cognitive complexity, while task performance under the simple‒complex condition displayed a U-shaped pattern in writing fluency, demonstrating the multidimensional, rather than strictly linear, nature of written production even over a short-term sequence (Malicka, 2014). Furthermore, the participants produced more words and demonstrated greater fluency in the simple-to-complex order than in isolation. Increased fluency in the simple task with planning corroborate the findings of previous studies (Farahani & Meraji, 2011; Levkina & Gilabert, 2012; Rahimi, 2018; Rostamian et al., 2018) and support Robinson's (2005) claim that the resource-directing factor's role (e.g., planning time) is to facilitate production under time pressure, assist access to learners' existing L2 knowledge, promote automaticity of the interlanguage system, and build fluency. Moreover, these results confirm the prediction of the Cognition Hypothesis that simpler tasks will improve production fluency (Baralt, Gilabert, & Robinson, 2014). Additionally, it is necessary to compare our results with the findings of Allaw and McDonough's (2019) study to better understand the short- and long-term effects of task sequencing on L2 writing fluency. While our results reveal a non-linear pattern regarding writing fluency over a short-term sequence, Allaw and McDonough (2019) reported that writing fluency is promoted as the result of increasing cognitive task complexity in the long term. Although the two studies manipulated tasks along resource-directing and resource-dispersing factors and employed the same measure to gauge L2 writing fluency, they differ from each other in terms of (1) the type of written task, (2) the types of resource-directing and resource-dispersing factors manipulated, (3) the explicit instructions participants received, and (4) the time intervals between the simple, less complex, and complex tasks. These differences can possibly account for diverging results between the two studies.

Implications and Future Directions

The current study offers theoretical and pedagogical implications for TBLT researchers and L2 teachers. Theoretically, our findings partially support the predictions of the Cognition Hypothesis and the SSARC model, revealing the progressive and variable nature of the effects of task sequencing when manipulated along resource-directing and resource-dispersing factors on a short-term simple-to-complex sequence. Increases in task complexity along the resource-directing factor induced learners to surpass their L2 knowledge base to meet the demands of the task, thus increasing the complexity of L2 written production. Conversely, the resource-dispersing factors facilitated access to existing L2 knowledge, promoting automatization and fluency. To help learners achieve more balanced CALF performance in writing, increasing these factors alone would not necessarily extend learners' L2 knowledge base and increase task complexity unless these factors are specifically integrated into simpler tasks in the sequence. It should be noted only a shorter sequence was tested and more research is warranted to examine how specific combinations of resource-directing and resource-dispersing factors could be integrated into a longer sequence to be used for a classroom syllabus. In line with the SSARC model, our findings also suggest that simple tasks designed for stabilization and complex tasks aimed at automatization should be implemented before subsequent restructuring and complexification can occur. This simple-to-complex sequence can provide learners with more scaffolded opportunities to practice and consolidate newly learned lexical and grammatical structures in their writing, promoting complexification of their production. However, this claim still requires more fine-grained evidence to flesh out the proposals of the SSARC model in the long term and disclose what happens after the processes of restructuring and complexification have occurred.

Pedagogically, the SSARC model associated with the Cognition Hypothesis offers several useful implications for classroom practice and syllabus design. Following the SSARC model, teachers can design lesson plans including a set of tasks with varying levels of cognitive complexity and provide multiple practice opportunities for learners to perform written tasks while benefiting from extra preparatory options such as pre-task planning time. At the pre-task stage, L2 learners with varying writing proficiency levels can experience less pressure and gain mental and linguistic preparedness for subsequent tasks due to scaffolding opportunities, improving focus on the content of their writing and increasing their speed of production. In the following stage, teachers can prompt learners to repeat the same task to facilitate greater autonomy, gradually increasing their independent performance as extra preparation is removed. Thus, learners may be more involved in rehearsal, the extension of their L2 repertoire, and development of complex structures, gaining prerequisite readiness for performing more sophisticated written tasks. In the post-task stage of writing, teachers can design more sophisticated resource-directing tasks, which stretch learners' linguistic resources, encourage them to face the challenges of performing these complex tasks, increase their ability to take risks, and push their interlanguage to its limits to use more innovative language. Teachers can also adjust the complexity level of tasks by manipulating different resource-directing and resource-dispersing factors; however, such task sequencing decisions should be made based on learners' needs, considering their background knowledge, proficiency levels, motivation, and readiness. In addition, teachers are cautioned against designing a task sequence with

different cognitive complexity levels in a vacuum. They should ensure that tasks with different instructional demands match relevant areas of their syllabus and promote learners' existing L2 knowledge over time. Finally, teachers with limited teaching experience may struggle to understand and interpret research-supported ideas, manipulate and sequence tasks, and regularly and effectively implement them in classroom settings. To solve this problem, teachers should receive ongoing training, support, and hands-on practice to understand the principles of the Triadic Componential Framework and the SSARC model, make sound task-sequencing decisions, and effectively operationalize classroom tasks with some degree of confidence.

This study has several limitations which require more consideration for future studies. As task sequencing was operationalized along both resource-directing and resource-dispersing factors, increasing the number of elements or the length of pre-task planning time could lead to different task sequencing effects on the CALF of L2 written production. Additionally, this study examined the effects of two task complexity factors (± number of task elements and ± planning time) on L2 writing. Future studies can extend this line of research by investigating the effects of manipulating other resource-directing and resource-dispersing factors on L2 written production. Furthermore, only subjective techniques were employed to verify the assumptions regarding the impacts of task manipulations on task complexity. Future studies could use objective techniques such as eye-tracking or dual task methodology to validate task difficulty assumptions in the L2 writing context. Finally, this study recruited participants at the upper-intermediate proficiency level, but more significant results could be obtained if advanced-level participants are examined to discover any relationships between proficiency and task sequencing effects.

Correspondence should be addressed to Mahmoud Abdi Tabari.
Email: ma4xf@virginia.edu

References

Abdi Tabari, M. (2020). Differential effects of strategic planning and task structure on L2 writing outcomes. *Reading & Writing Quarterly*, *36*(4), 320–338. https://doi.org/10.1080/10573569.2019.1637310

Abrams, Z. I. (2019). The effects of integrated writing on linguistic complexity in L2 writing and task-complexity. *System, 81*, 110–121. https://doi.org/10.1016/j.system.2019.01.009

Allaw, E. & McDonough. K. (2019). The effect of task sequencing on second language written lexical complexity, accuracy, and fluency. *System, 85*, 102014. https://doi.org/10.1016/j.system.2019.06.008

Baralt, M. (2014). Task complexity and task sequencing in traditional versus online language classes. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 95–122). Bloomsbury.

Baralt, M., Gilabert, R., & Robinson, P. (2014). An introduction to theory and research in task sequencing and instructed second language learning. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 1–37). Bloomsbury.

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

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

Farahani, A. A. K., & Meraji, S. Z. (2011). Cognitive task complexity and L2 narrative writing performance. *Journal of Language Teaching and Research, 2*(2), 445-456. https://doi.org/10.4304/jltr.2.2.445-456

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375. https://doi.org/10.1093/applin/21.3.354

Frear, M.W., & Bitchener, J. (2015) The effects of cognitive task complexity on writing complexity. *Journal of Second Language Writing, 30*, 45-57. https://doi.org/10.1016/j.jslw.2015.08.009

Hunt, K. (1970). *Syntactic maturity in schoolchildren and adults*. University of Chicago Press.

Jackson, D. O., & Suethanapornkul, S. (2013). The cognition hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning, 63*(2), 330–367. https://doi.org/10.1111/lang.12008

Johnson, M. D. (2017). Cognitive task complexity and L2 written syntactic complexity, lexical complexity, accuracy, and fluency: A research synthesis and meta-analysis. *Journal of Second Language Writing, 37*, 13–38. https://doi.org/10.1016/j.jslw.2017.06.001

Kim, Y., & Payant, C. (2014). A pedagogical proposal for task sequencing: An exploration of task repetition and task complexity on learning opportunities. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 151–177). Bloomsbury.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing, 20*(2), 148–161. https://doi.org/10.1016/j.jslw.2011.02.001

Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *IRAL: International Review of Applied Linguistics in Language Teaching, 45*(3), 261–284. https://doi.org/10.1515/iral.2007.012

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. https://doi.org/10.1016/j.jslw.2007.08.003

Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking: The effect of mode. In P. Robinson (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 91–104). John Benjamins. https://doi.org/10.1075/tblt.2

Kuiken, F., & Vedder, I. (2012). Syntactic complexity, lexical variation and accuracy as a function of task complexity and proficiency level in L2 writing and speaking. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143–170). John Benjamins. https://doi.org/10.1075/lllt.32

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757–786. https://doi.org/10.1002/tesq.194

Lambert, C., & Robinson, P. (2014). Learning to perform narrative tasks: A semester-long classroom study of L2 task sequencing effects. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 207–230). Bloomsbury.

Lee, J. (2018). *The interactive effects of task complexity, task condition, and cognitive individual differences on L2 writing* [Unpublished doctoral dissertation], University of Maryland.

Levkina, M., & Gilabert, R. (2014). Task sequencing in the L2 development of spatial expressions. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 37–70). Bloomsbury.

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

Malicka, A. (2014). The role of task sequencing in L2 oral production. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 71–95). Bloomsbury.

Malicka, A. (2018). The role of task sequencing in fluency, accuracy, and complexity: Investigating the SSARC model of pedagogic task sequencing. *Language Teaching Research*, 1-24. https://doi.org/10.1177%2F1362168818813668

Malicka, A., & Levkina, M. (2012). Measuring task complexity: Does L2 proficiency matter? In A. Shehadeh & C.A. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 43–67). John Benjamins.

Malicka, A., & Sasayama, S. (2017). *The importance of learning from accumulated knowledge: Findings from a research synthesis on task complexity*. Unpublished paper presented at the 7th International Conference on Task-Based Language Teaching (TBLT), University of Barcelona, Barcelona, Spain.

Manchón, R. M. (2014). The internal dimension of tasks: The interaction between task factors and learner factors in bringing about learning through writing. In H. Byrnes & R.M. Manchón (Eds.), *Task-based language learning: Insights from and for L2 writing* (pp. 27–52). John Benjamins. https://doi.org/10.1075/tblt.4

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

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods, 42*(2), 381–392. https://doi.org/10.3758/BRM.42.2.381

McNamara, D., Graesser, A.C., McCarthy, P.M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Michel, M., Révész, A., Lu, X., Kourtali, N., Lee, M., & Borges, L. (2020). Investigating L2 writing processes across independent and integrated tasks: A mixed-methods

study. *Second Language Research, 36*(3), 307-334. https://doi.org/10.1177/0267658320915501

Michel, M., & Smith, B. (2018). Measuring Lexical Alignment During L2 Chat Interaction: An Eye-Tracking Study. In S.M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in Second Language Acquisition* (pp. 244-267). Routledge. https://doi.org/10.4324/9781315399027

Norris, J. M. (2010). Understanding instructed SLA: Constructs, contexts, and consequences. Plenary address delivered at the annual conference of the European Second Language Association (EUROSLA).

Norris, J.M., Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*, 555–578. https://doi.org/10.1093/applin/amp044

Ong, J., & Zhang, L. J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing, 19*(4), 218–233. https://doi.org/10.1016/j.jslw.2010.10.003

Plonsky, L., & Oswald, F.L. (2014) How big is 'big'? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. https://doi.org/10.1111/lang.12079

Rahimi, M. (2018). *Task complexity, affective factors, and pre-task planning: Effects on L2 writing production* [Unpublished doctoral dissertation], University of Auckland.

Rahimi, M. (2019) Effects of increasing the degree of reasoning and the number of elements on L2 argumentative writing. *Language Teaching Research, 23*(5), 633-654. https://doi.org/10.1177%2F1362168818761465

R Development Core Team. (2019). *R* (version 3.6.1). The R Foundation. https://www.r-project.org/

Révész, A. (2014). Towards a fuller assessment of cognitive models of task-based learning: Investigating task-generated cognitive demands and processes. *Applied Linguistics, 35*, 87–92. https://doi.org/10.1093/applin/amt039

Révész, A., Kourtali, N. E., & Mazgutova, D. (2016). Effects of task complexity on L2 writing behaviors and linguistic complexity. *Language Learning, 67*, 208–241. https://doi.org/10.1111/lang.12205

Révész, A., Michel, M., & Gilabert, R. (2015). Measuring cognitive task demands using dual-task methodology, subjective self-rating, and expert judgments. *Studies in Second Language Acquisition, 38*, 703–37. https://doi.org/10.1017/S0272263115000339

Révész, A., Sachs, R., & Hama, M. (2014). The effects of task complexity and input frequency on the acquisition of the past counterfactual construction through recasts. *Language Learning, 64*, 615–650. https://doi.org/10.1111/lang.12061

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics, 22*, 27–57. https://doi.org/10.1093/applin/22.1.27

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

Robinson, P. (2007). Criteria for classifying and sequencing pedagogic tasks. In M. del Pilar Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 7–26). Multilingual Matters.

Robinson, P. (2010). Situating and distributing cognition across task demands: The
        SSARC model of pedagogic task sequencing. In M. Pütz & L. Sicola (Eds.),
        Cognitive processing in second language acquisition: Inside the learner's mind
        (pp. 243–268). John Benjamins. https://doi.org/10.1075/tblt.8
Robinson, P. (2015). The Cognition Hypothesis, second language task demands, and the
        SSARC model of pedagogic task sequencing. In M. Bygate (Ed.), *Domains and
        directions in the development of TBLT: A decade of plenaries from the
        international conference* (pp. 87–122). John Benjamins.
Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second
        language learning and performance. *International Review of Applied Linguistics,
        45*(3), 161–284. https://doi.org/10.1515/iral.2007.007
Rostamian, M., Fazilatfar, A. M., & Jabbari A. A. (2018) The effect of planning time on
        cognitive processes, monitoring behavior, and quality of L2 writing. *Language
        Teaching Research*, *22*(4), 418-438.
        https://doi.org/10.1177%2F1362168817699239
Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of
        cognitive task complexity. *The Modern Language Journal, 100*(1), 231-254.
        https://doi.org/10.1111/modl.12313
Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and
        second language instruction* (pp. 183–205). Cambridge University.
Tavakoli, P. (2014). Storyline complexity and syntactic complexity in writing and speaking
        tasks. In H., Byrnes, H. & R. Manchón (Eds.), *Task-based language learning:
        Insights from and for L2 writing* (pp. 163–191). John Benjamins.