Article

LANGUAGE TEACHING RESEARCH

Language Teaching Research
1–24

DOI: 10.1177/1362168818813668
journals.sagepub.com/home/ltr
SAGE

The role of task sequencing in fluency, accuracy, and complexity: Investigating the SSARC model of pedagogic task sequencing

Aleksandra Malicka
Universitat Oberta de Catalunya, Spain

Abstract

This study set out to test the theoretical premise of the SSARC model of pedagogic task sequencing, which postulates that tasks should be sequenced for learners from cognitively simple to complex. This experiment compared the performance of three tasks differing in cognitive complexity in a simple–complex sequence versus in the absence of any other tasks. There were two groups in the study: (1) participants who performed the three tasks in the simple–complex sequence, and (2) participants who performed either the simple, the complex, or the most complex task. The participants' speech was analysed using fluency, accuracy, and complexity measures. The results indicate that simple–complex sequencing led to a higher speech rate, greater dysfluency, enhanced accuracy, and greater structural complexity, as compared to individual task performance. The results are discussed in terms of the SSARC model and pedagogical implications of the findings are presented.

Keywords

oral production, SSARC model, task-based language teaching, task design, task sequencing

I Introduction

For several decades now, the construct of task complexity has been at the forefront of conceptual and empirical investigations in the domain of task-based language teaching (TBLT). Grounded in the cognitive research paradigm, a vast amount of scholarly interest has focused on attempting to discover whether universal task design characteristics exist which influence learners' linguistic behaviours in systematic, predictable ways.

Corresponding author:
Aleksandra Malicka, Department of Arts and Humanities, Universitat Oberta de Catalunya, Avinguda Tibidabo, 47, Barcelona, 08035, Spain.
Email: amalicka@uoc.edu

Recent years have witnessed a surge in studies which set out to investigate the effects of task complexity in relation to three broad outcomes: production, typically conceived of as fluency, accuracy, and lexical and structural complexity (e.g. Ahmadian, 2012; Benson, 2015; Hsu, 2015; Ishikawa, 2011; Kormos & Trebits, 2011; Kuiken & Vedder, 2011; Lambert & Robinson, 2014; Michel, 2011), acquisition (the development of specific linguistic forms; e.g. Baralt, 2013, 2014; Kim & Taguchi, 2015; Kim & Tracy-Ventura, 2011; Kim, Payant, & Pearson, 2015; Levkina & Gilabert, 2014), and interaction (e.g. Kim & Payant, 2014; Nuevo, Adams, & Ross-Feldman, 2011). A large portion of this body of empirical work has been informed by two robust theoretical frameworks: Robinson's (2001a, 2003) Cognition Hypothesis and Skehan's (1996, 1998) Trade-off Hypothesis. Within the scope of these frameworks, the central line of scientific interest has revolved around comparing the effects that designed-to-be-cognitively simple tasks, versus their complex counterparts, have on learners' linguistic behaviours. More specifically, empirical research has set out to explore such issues as:

1. What is the effect of manipulating task complexity on fluency, accuracy, lexical and structural complexity of learners' production in a second or foreign language?
2. How does the exposure to cognitively simple vs. complex tasks affect the development of specific linguistic forms, for example English progressive, Spanish past subjunctive, or the acquisition of specific vocabulary items?
3. Does engaging in simple vs. complex tasks lead to different patterns of occurrence of interactional episodes, such as negotiation of meaning, confirmation checks, clarification requests, etc.?
4. What role do individual differences, such as learner's proficiency level or working memory capacity, play in language performance, learning, and interaction?

While providing definitive answers to these questions is beyond the scope of any individual empirical endeavor, recent syntheses of research in the TBLT domain have started to discern patterns across studies, advancing our understanding of the effects that different types of tasks and variables have on learners' linguistic behaviours. Regarding specifically the research agenda which focuses on the effects of task complexity on production, in a meta-analysis of studies investigating the Cognition Hypothesis, Jackson and Suethanapornkul (2013) found that performing complex monologic oral tasks leads to enhanced accuracy and greater lexical complexity, but less fluency and structural complexity, compared to performance on simpler counterpart tasks. However, the results were based on a sample of only 9 studies, and the detected effect sizes were small. According to the preliminary results of a comprehensive research synthesis and meta-analysis of cognitive task complexity (Malicka & Sasayama, 2017; Sasayama, Malicka, & Norris, 2015) the following claims can be formulated:

1. Simple tasks along variables such as planning time or task repetition are associated with positive effects on all production dimensions (with substantial effects found in the case of accuracy and structural complexity, but smaller in the case of fluency and lexical complexity).

2. Complex tasks along the variable here-and-now/there-and-then lead to greater syntactic complexity.
3. Complex tasks manipulated along the variables number of elements and reasoning, generate more lexical complexity.

Substantial conceptual advances in the area of task complexity, such as the development of the Cognition Hypothesis or the Trade-off Hypothesis, together with a sizeable body of empirical work carried out thus far, bring us closer to making reasonably safe claims about the effects of task design on language performance, and further our understanding of the complex processes involved in the production and acquisition of second languages. Still, despite a sustained scholarly interest in a host of issues related to task design (e.g. the impact of variables, task types, and task conditions on performance, interaction, and acquisition), and despite recent efforts to understand the broad domain by systematically synthesizing research findings, the issue of how to optimally sequence pedagogic tasks for learners represents a major conceptual challenge, and it remains an open empirical question. While research into the effects of task complexity on different outcomes undoubtedly informs the theorizing about task sequencing, it is fundamentally unclear on the basis of what criteria tasks should be sequenced for learners so as to maximize opportunities for learning to occur, and to create an optimal learning environment. As stated by Baralt, Gilabert, and Robinson (2014, p. 1), 'To date …, there is still no widely agreed-upon set of criteria that can be used to grade and sequence tasks.' At the same time, establishing theoretically driven, pedagogically sound and researchable criteria for sequencing tasks is of vital importance both for learners' immediate (production) and long-term (acquisition) goals in a second language.

The most recent and most robust proposal of task sequencing thus far, Robinson's (2010) SSARC model of pedagogic task sequencing, associated with the Cognition Hypothesis, makes two fundamental claims: (1) sequencing decisions should be based solely upon cognitive factors, and (2) tasks should be sequenced in the order from simple to complex. Building on this theoretical premise, and on the findings from previous research, the current study aims to contribute to the scholarly debate on the role of sequencing by comparing performance on tasks sequenced from simple to complex versus performance on individual tasks, a methodological approach not investigated thus far. In order to contextualize the empirical research reported in this study, Section II summarizes the efforts towards conceptualizing task sequencing prior to the development of the SSARC model.

II Early conceptualizations of task complexity and task sequencing

Theorizing about the notions of task complexity and sequencing dates back to early 1980s. One of the first classifications was offered by Brown, Malmkjaer, and Williams (1984), who suggested basing sequencing criteria on the level of abstractness inherent in a task. They proposed a hierarchy in which abstract notions (argumentation or justification) were considered more difficult than dynamic relationships (e.g. referring to actions, activities or processes). Dynamic relationships, in turn, were more difficult than static relationships

(relationships among objects). The number of objects – and the difficulty of the relationship between them – were the factors determining relative difficulty. Finally, 'tasks which require the speaker to communicate abstract notions, for instance in argument or justifications, are more difficult again' (Brown et al., 1984, p. 51).

Along similar lines, one of the early criteria for sequencing was based on task types. Prabhu (1987) in the Bangalore Project assessed different task types potentially as representing different levels of challenge for the speaker. He proposed a hierarchy in which opinion-gap tasks presented a greater challenge than reasoning-gap tasks, and reasoning gap tasks presented a greater challenge than information-gap tasks. The lowest level in this hierarchy was attributed to simple information transmission, followed by information transmission requiring inferencing and deduction. The greatest challenge was attributed to those tasks which dealt with preferences, attitudes, feelings, and beliefs.

The late 1980s witnessed a conceptual advancement in theorizing about the notion of task complexity and sequencing, as evidenced in the emergence of somewhat more tangible criteria. Candlin (1987) identified six parameters along which sequencing decisions can be taken. Cognitive load referred to clear, logical sequence versus a lack of it, individual versus multiple characters, and individual versus multiple actions. Communicative stress involved such factors as topic familiarity (known vs. unknown) and number of interlocutors (one or few vs. many), and a series of interlocutors' characteristics: their level of communicative competence, their familiarity with the subject matter, and the extent to which a task followed a clear organization. Code complexity and interpretive density were concerned with the complexity of texts used in tasks, in linguistic and argumentation terms. Content continuity referred to how much a pedagogic task resembled a real-world target task. Process continuity was related to learners' choice about sequencing tasks. Finally, particularity and generalizability referred to the relative familiarity or unfamiliarity of a situation.

In a similar vein, Brindley (1987) proposed a three-way distinction between the following groups of factors: learner factors (confidence, motivation, prior learning experience, ability to learn at the pace required, possession of necessary language skills, and relevant culture knowledge), task factors (degree of cognitive complexity, number of steps, amount of context support, and amount of time provided), and text factors (length and clarity of text).

Similar to Brindley's approach is that of Nunan (1989), who put forward a set of criteria including input factors, which encompassed such aspects as grammatical complexity of the input (simple sentences vs. sentences containing subordination), length of a text, propositional density, the amount of low-frequency vocabulary, the speed of spoken texts, the number of speakers involved, the explicitness of the information, and discourse structure. Learner factors were those which the learner brought to the task: background knowledge, linguistic knowledge, confidence, interest, motivation, and observed ability in language skills. Finally, procedural factors were 'operations that the learners are required to perform on input data' (Nunan, 2004, p. 122), and they included such components as relevance (relative relevance of the task to the learner), complexity (steps involved, complexity of instructions, cognitive demands, amount of information), or context provided (required prior knowledge, preliminary activity), among others.

What these early attempts at establishing sequencing criteria had in common was that they conflated factors clearly belonging to different categories, some perhaps being well suited to deliberate manipulations in task design, such as cognitive load (e.g. changing the number of characters or parallel actions, code complexity, text factors, or procedural factors), and others being external sources affecting overall complexity of tasks. While cognitive load (e.g. task factors or input factors) potentially lends itself to being controlled in task design, and it can therefore serve as a sequencing criterion, it is unclear how those factors which are not predictable in advance of instruction, such as individual differences, can function as criteria upon which to sequence tasks. Clearly, all factors influence performance in unique ways, but any attempts at theorizing about their impact should disentangle the factors subject to manipulation from non-manipulable and non-controllable external influences. This is an area in which the model investigated in this article – the SSARC model of pedagogic task sequencing – presents a conceptual advancement over early sequencing proposals.

III SSARC model of pedagogic task sequencing

According to Robinson (2001b, p. 27), ‘The development of theoretically motivated, empirically substantiable, and pedagogically feasible sequencing criteria has long been acknowledged as a major goal of research aimed at operationalizing task-based approaches to syllabus design.’ Robinson postulated the development of a sequencing model which should meet three criteria: (1) being well-grounded in theory, (2) being researchable, and (3) lending itself to pedagogic interventions. The SSARC model he put forward meets these principles, and builds on Robinson’s previous conceptual work: the Cognition Hypothesis (CH) and the Triadic Componential Framework (TCF; Robinson, 2005, 2007).

Grounded in the Multiple Attentional Resource Model (Wickens, 2002, 2007), the CH draws on the idea that speakers of a second language have at their disposal multiple attentional resources, rather than a single volume of attention which runs out of resources. This idea is fundamental to the subsequent theorizing about what happens with the different dimensions of production as task demands increase: given the multiplicity of attentional resources, the dimensions of linguistic complexity and accuracy can be processed in memory and attended to at the same time, with a possible simultaneous detriment to fluency. The Triadic Componential Framework (Robinson, 2005, 2007), associated with the CH, systematically classifies different manipulable and non-manipulable task variables, and it is made up of three building blocks: Task Complexity (cognitive factors), Task Condition (interactive factors), and Task Difficulty (learner factors). Of particular interest to the SSARC model, and to the current study, is the cognitive group of factors. In the TCF, this group of factors is further divided into two subgroups of variables: resource-directing (e.g. reasoning demands) and resource-dispersing (e.g. planning time), each of which has a unique role in performance and learning: while the former gear learners’ attention towards linguistic aspects of tasks, the latter disperse learners’ attention over many non-linguistic aspects of tasks. Consequently, Robinson (2001a) predicts that complex interactive and monologic resource-directing tasks lead to greater accuracy (and in the case of monologic tasks, complexity), but less fluency than simpler tasks. By contrast, on resource-dispersing dimensions, more complex interactive and monologic tasks lead to less accuracy, fluency, and complexity.

Crucial to understanding Robinson's stance on sequencing is precisely the isolation of cognitive factors from other types of factors or sources of influence: interactive and learner factors. Robinson postulates that task complexity, and its two building blocks – resource-directing and resource-dispersing variables – should be the sole basis for sequencing criteria, which constitutes the first task sequencing principle in the SSARC model (Robinson, 2010). Robinson supports his claim by stating that such sequencing helps 'ensure deep semantic processing (Craik & Lockhart, 1972; Hulstijn, 2001, 2003) rehearsal in memory (Robinson, 2003) and elaboration and successful transfer of the particular "schema" for interactive or monologic task performance to real-world contexts of use (Schank, 1999; Schank & Abelson, 1977)' (Robinson, 2010, p. 247). What such sequencing means in practical terms is that tasks should be sequenced along such variables as ±reasoning demands, ±here and now, ±elements ('resource-directing dimensions'), or ±planning time, ±single task, ±few steps ('resource-dispersing dimensions'), but not on the basis of, for example, participation variables (e.g. ±open solution, ±few participants), or participant variables (±same proficiency, ±same gender). Robinson claims that 'task sequencing is done by designing and having learners perform tasks simple on all the relevant parameters of task demands first, and then gradually increasing their cognitive complexity on subsequent versions' (Robinson, 2010, p. 246–247). Such ostensibly facilitative function of sequencing tasks from simple to complex on the basis of cognitive criteria is what this article aims to explore.

The second sequencing principle deals with the order in which different task complexity factors should be manipulated. Robinson proposes three stages in implementing a sequence of tasks[1]:

- Step 1: tasks simple on resource-directing and resource-dispersing dimensions ('SS' for simple, stable attractor state): SS = i × e ('s'rdisp) + ('s'rdir) ^{n}
- Step 2: tasks cognitively demanding on resource-dispersing dimensions ('A' for 'automatization'): A = i × e ('c'rdisp) + ('s'rdir) ^{n}
- Step 3: tasks cognitively demanding on both resource-directing and resource-dispersing dimensions: ('RC' for restructuring and complexifying interlanguage): RC = i × e ('c'rdisp) + ('c'rdir) ^{n}.

Following this principle, for instance a narrative task should first be performed with the provision of planning time (task simple on resource-dispersing dimension) and in the here-and-now condition (task simple on resource-directing dimension). When a task is simple on both types of dimensions, task performance relies on the simple, stable ('SS') state of current second language (L2) proficiency. On a more complex version, the same task should be narrated in the present (task simple on the resource-directing dimension: here-and-now/there-and-then), without provision of planning time (task complex on the resource-dispersing dimension). This cognitively more demanding task version fosters consolidation, access to, and automatization ('A') of the learner's current interlanguage system. Finally, the most complex task version is narrated in the past, with no provision of planning time (task complex on both resource-directing and resource-dispersing dimensions). Introducing maximum cognitive complexity, instantiated by making the task complex on both resource-directing and resource-dispersing variables, leads to interlanguage restructuring, and to the development of new form–function/concept mappings. It also triggers maximum complexity (C) of the interlanguage and its destabilization. A learner's attention is divided over non-linguistic aspects of the task (triggered by resource-dispersing dimensions), and simultaneously resource-directing dimensions direct the attention to the task's linguistic aspects. Robinson argues that increasing cognitive complexity along resource-directing dimensions triggers attention paid to form–function mappings, and therefore potentially promotes interlanguage development, while complexity increased along resource-dispersing dimensions 'promotes increasing automatic access to current linguistic resources' (Robinson, 2010, p. 247).

Robinson's model is conceptually more advanced than previous proposals in three aspects. First, unlike early theorizing on sequencing, the SSARC model is theoretically-driven. It is informed by the Cognition Hypothesis, which draws on the conceptual and empirical work in the area of resource theory. Second, both the task taxonomy in the Triadic Componential Framework and in the SSARC model separate cognitive factors from other kinds of factors (task condition and learner factors), and suggest scaling up the complexity of tasks only on the basis of cognitive complexity. Third, as could be observed in previous proposals, factors clearly belonging to different categories were conflated, with some being manipulable and others not. Others, on the other hand, were external sources alleviating or augmenting overall cognitive complexity of tasks. Clearly, separating these factors in the SSARC model is conceptually more advanced as it distinguishes the controllable, manipulable internal task features from other non-controllable and non-manipulable external sources of influence.

IV Review of selected studies on task sequencing

Several authors set out to test the fundamental theoretical claim of the SSARC model about the ostensibly facilitative role simple–complex sequencing has for performance and learning in a second language. An approach taken in a number of studies consisted in comparing this sequencing order with alternative sequencing orders (complex–simple and randomized). The role of sequencing was measured in relation to production (Malicka, 2014), acquisition (Baralt, 2014; Levkina & Gilabert, 2014), and interaction (Kim & Payant, 2014).

Baralt (2014) compared the effects of the simple–complex sequence versus other sequences (CCS, CSC, and SCS) with a story-retell task in traditional, face-to-face versus computerized online settings. This study explored how different sequencing orders influence the acquisition of the Spanish past subjunctive. While sequencing was shown to play a role in the acquisition of this structure in the traditional classroom, there were no instances of the Spanish past subjunctive in the online modality, irrespective of the sequencing order.

In a study with a similar design, Levkina and Gilabert (2014) tested the SSARC model by comparing three sequences of tasks (simple–complex, complex–simple, and randomized). They enquired about these sequences' effects on the acquisition of spatial expressions. The participants were required to perform three information-gap tasks in which they played the role of a delivery person and, following the instructions of the client, they had to furnish different spaces in an apartment. The acquisition of the target structure was measured on a productive task and on a receptive vocabulary test. In the immediate post-test, the participants in all three sequencing conditions improved their vocabulary knowledge, disconfirming the predictions of the SSARC model about the alleged advantage of simple–complex sequencing over alternative orders. Concerning long-term gains, the participants in the simple–complex condition produced a significantly higher number of target items in the delayed post-test administered two weeks later, as compared to the immediate post-test.

In a study with a similar design, but focused on language production, Malicka (2014) enquired about the benefits of simple–complex sequencing order (simple, ‘S’; complex, ‘C’; and most complex, ‘+C’) as opposed to randomized sequencing, which was operationalized as five different sequences: (1) S+CC, (2) CS+C, (3) C+CS, (4) +CCS, and (5) +CSC. In each condition, the participants performed three tasks whose cognitive complexity was manipulated along the variables ±elements and ±reasoning demands. The differences between the three tasks, as measured by global fluency, accuracy, and complexity indices, turned out to be statistically significant, indicating a strong influence of cognitive complexity on performance. However, no interaction was found between task complexity and sequencing tasks from simple to complex; that is, the cognitive complexity of tasks, and not the sequencing order, was what led to short-term qualitative changes in performance, a result similar to that found by Baralt (2014).

Kim and Payant (2014) explored the effects of task repetition and increasing cognitive complexity on the occurrence of language-related episodes (LREs). The study involved two types of repetition: task repetition (repeating the same task), and procedural repetition (repeating the same procedure, and a comparable, but not identical task). First language (L1) Korean learners of English were assigned to one of these four experimental conditions: (1) simple task content/procedural repetition, (2) simple task procedural repetition, (3) complex task content/procedural repetition, and (4) complex task procedural repetition. The participants in the procedural repetition condition presented a higher number of LREs than those who repeated the same task. Furthermore, the production of LREs was not influenced by task complexity, that is, performing the complex task did not lead to a greater number of LREs than performing the simple task.

In the body of empirical work carried out so far, juxtaposing simple–complex sequencing with alternative orders has been the default experimental design. The findings reported here allow to draw the conclusion that overall studies have failed to provide support for the claim that simple–complex sequencing presents advantages over alternative sequencing orders. In order to further our understanding of task sequencing, contribute to its conceptualization, and to offer an alternative approach to testing the premises of the SSARC model, the current study put to test sequencing tasks from simple to complex and compared tasks sequenced from simple to complex with individual, isolated task performance (baseline data). The rationale for such an approach lies in the need to discover how performance of tasks in a sequence from simple to complex compares to the same tasks performed on their own. The fundamental difference between the design offered in the current study and previous literature is that in the experiment reported here, on one hand a group of participants performed a sequence of three tasks differing in cognitive complexity, and on the other hand the same tasks were performed by three different groups of participants, each of which performed either the simple, the complex, or the most complex task version. Therefore, unlike previous research, this study aims to explore the ostensibly facilitative role of simple–complex sequencing in relation to baseline data, and not in relation to alternative sequencing orders.

V Methods

1 Research questions

Motivated by the above described rationale, the research question this article addresses is as follows: Does performing a simple, complex, and the most complex task in the sequence from simple to complex, versus performing the same tasks in isolation, leads to qualitatively different output in terms of fluency, accuracy, and complexity?

2 Participants

The experiment was done in an English for specific purposes context, in a curriculum focused on tourism. The subjects (n = 87) were L1 Spanish and Catalan learners of English as a foreign language. They were undergraduate students at a college of tourism. Their age ranged from 19 to 31 years (M = 21; SD = 2.38). Twenty-four males and 63 females took part in the experiment. English was an obligatory subject on the curriculum; however, the administered tasks were not part of the curriculum. The participants' proficiency was measured by means of Oxford Placement Test (OPT), extensively used in previous research (e.g. Ahmadian, 2012; Murphy & Roca de Larios, 2010; Tavakoli & Foster, 2008). It is a multiple-choice test with 60 items and it is targeted at lexis and syntax. This test was administered during class time several weeks before the data collection took place. (Section V.6 reports the results of the proficiency test.)

3 Treatment tasks: Context

There were three treatment tasks in the study. The participants were asked to imagine that they had just started working as hotel receptionists and the hotel they worked for was famous for its personal touch when dealing with its clients. According to the hotel's policy, all new employees had to engage in training sessions in order to practice their skills and become more self-confident when it comes to dealing with the clients. The designed tasks represented three situations in which the receptionist had to solve a problem. The receptionist's role was to relocate the clients to different rooms or hotels. Each of the three tasks included profiles of the clients and the room or hotel options to choose from. In the simple task, the participant-receptionist had to describe the different options offered by the hotel, and in the other two tasks (the complex and the most complex task), they needed to make a decision about which room or hotel option best matched each of the clients on the basis of the characteristics of both the clients and the room and hotel options. The complex and the most complex task were therefore based on a mismatch between the clients' requests and the options the hotel could offer. There was not a single perfect solution the receptionist could offer, but rather several imperfect ones, which rendered the task open in terms of outcome.

4 Treatment tasks: Cognitive complexity operationalization

The theoretical basis for the development of tasks was Robinson's Cognition Hypothesis and the Triadic Componential Framework. Two variables from this framework were manipulated in task design: ±number of elements and ±reasoning demands. 'Reasoning demands' is the amount of online computation required to perform a task. Increased cognitive complexity along this variable involves augmenting the number of mental operations so as to pose a challenge so substantial on the learner's attentional resources that it leads to the restructuring of the L2 system. The tasks were designed around the build-up of mental operations needed for successful task completion, and these were conceived of as thought processes such as reasoning from hotel options available to recommend a best choice for a customer; justifying room selections made by giving reasons; and apologizing with explanations and reasons for a state of affairs. Increased cognitive complexity was the function of the number of these operations and their simultaneous occurrence in a task.

'Number of elements' refers to the number of occurrences of a specific task component. This variable referred to client profiles and, depending on the task, hotel or room options. The number of the clients (3) and the number of room/hotel options (5) were kept constant across the three tasks. Linking this operationalization of variables to the SSARC model, when tasks are performed in a sequence, a simple task on both reasoning demands and number of elements plays a facilitative role of stabilizing the L2 system, a complex task automatizes access to L2 forms, and the most complex task restructures the L2 system in search of optimal ways to deal with high cognitive demands.

In the simple task, the participants were supposed to describe the room options available at the hotel to three clients, without taking into consideration any client characteristics; therefore, this task required only simple information transmission. The complex task was designed around the situation of overbooking, and the participants had to engage in the mental operations of apologizing, describing, and recommending an alternative option suitable to different clients. The number of mental operations therefore increased compared to the simple task, and the number of elements rose from none to many in the case of the pieces of information which build the client profile. The most complex task, also designed around the situation of overbooking, once again, required finding the best match for each client. There were multiple elements to be taken into consideration regarding the client profile, and one more mental operation in addition to those present in the complex task: justifying one's choice.[2] Task input was presented to the participants in L1 Spanish in order for the tasks to resemble real-life tasks, and in order to prevent the participants from merely repeating the lexis and structures encountered in the input in their productions.

5 Validating the assumptions of cognitive task complexity

As described above, the three tasks employed in this experiment were designed such that they represented different levels of cognitive challenge for the learners. In order to validate the assumptions about cognitive complexity in task design, a pilot study with 23

Table 1. Affective variables questionnaire: Means and standard deviations.

Measured construct	Simple task		Complex task		Most complex task	
	M	*SD*	*M*	*SD*	*M*	*SD*
Difficulty	3.35	1.36	4.3	1.27	5.8	1.82
Mental effort	3.48	1.5	4.43	1.19	6.1	2.02

Table 2. Time judgment task: Means and standard deviations.

Simple task		Complex task		Most complex task	
M	*SD*	*M*	*SD*	*M*	*SD*
73.6	74.61	56.21	60.17	54.94	82.47

Note. The means denote the difference, in seconds, between the estimated and the actual time it took the participants to perform the tasks.

participants was carried out prior to the main experiment. The participants first performed the three tasks,[3] after which a battery of three measures were put in place in order to validate the assumptions behind task complexity manipulation: affective variables questionnaire (AVQ), time judgment task (TJT), and stimulated recall (SR).

Right after performing the task, the participants were handed a copy of the AVQ and TJT (see Appendix 1). In the AVQ, they were required to mark their judgments of two constructs: difficulty and mental effort involved in performing the task, on a 9-point Likert scale. The results are displayed in Table 1. There was a gradual pattern of increase in perceived difficulty in the case of both constructs: ST was perceived as the least difficult task, CT as more difficult than the simple one, and +CT task as the most difficult one. The perception of difficulty between CT and +CT was greater than that between the ST and the CT. Also, the means obtained for mental effort were generally slightly higher than those obtained for difficulty.

In the time judgment task (Baralt, 2010; Malicka & Levkina, 2012; Révész, Michel, & Gilabert, 2012), the participants were asked to assess the time it took them to perform the task. Their estimations were subsequently compared with the actual time they took to perform the tasks. The prediction was that as cognitive complexity increases, the greater the expected mismatch should be between the real time and the estimated time. As can be observed in Table 2, the opposite trend to that expected was revealed: the greatest mismatch between the real and guessed time was detected in the case of the simple task, and the values obtained for the two other tasks were lower, with only a minor difference found between them. Therefore, the results of this measure did not corroborate those yielded by the AVQ. As can also be observed in all the tasks, the standard deviation values are bigger than the means, indicating a wide data dispersion in the case of this measure.

Finally, stimulated recall was done with four participants. After completing the AVQ and TJT, they were required to orally reflect in retrospect on cognitive complexity of tasks via two questions: (1) Did you think there were differences in difficulty between the tasks you performed? (2) Why did you think they differed in difficulty? In three out of four cases, the information provided by the participants matched the hypothesized differences in cognitive complexity. In their answers, the participants indicated that the more options they had to deal with (i.e. client and room or hotel characteristics), the more difficult the task was because they had to pay simultaneous attention to many aspects of the tasks. They also pointed to the fact that task difficulty increased when they had to perform several mental operations simultaneously (i.e. describing, apologizing, convincing, etc.). These insights matched the designed differences in cognitive complexity in both variables: number of elements and reasoning demands.

6 Experimental design

There were two groups in the study: (1) simple–complex sequencing (n = 30), and (2) individual task performance (baseline data) (n = 57). In the first group the three tasks were performed by the same participant in a subsequent fashion. Sequencing was operationalized as performing a sequence of three tasks in one sitting, at short, 1-minute intervals, with unlimited online planning time. All participants performed these tasks in accordance with the SSARC model of pedagogic task sequencing, that is, first they completed the simple task, then the complex one, and finally the most complex one.

In the individual task performance (ITP) group, three subgroups were involved. The participants in each of them performed one task of a specific cognitive complexity level, either the simple task (n = 18), the complex task (n = 19), or the +complex task (n = 20). Therefore, the difference between the two experimental conditions is that in the first group, the same subject performed three tasks, and in the second one, one subject performed only one task.

The results of the proficiency test (OPT) obtained by the different groups in the study were as follows (the maximum score was 60): M = 32.82, SD = 9.12 for the participants in the simple–complex condition; M = 28.88, SD = 7.67 for the participants who performed the simple task; M = 32.36, SD = 6.63 for the participants who performed the complex task; M = 33.78, SD = 9.54 for the participants who performed the most complex task. An independent samples t-test did not yield statistically significant differences between the groups. It was concluded that they were homogenous in terms of their L2 level competence.

7 Data collection procedure

Data collection took place during regular class time with every student performing the task individually one-on-one with the researcher. Before data collection, the context and general instructions to all tasks were read out aloud by the researcher in Spanish. The participants were told that they would participate in a training session for hotel receptionists, and that they would perform one or several tasks in which they would have to solve a problem at a hotel reception. The participants had one minute to familiarize themselves with the instructions before starting the task. It was sufficient to get a general idea of what the task was about, and scan the instructions. The participant was sitting facing the researcher, at the teacher's desk. First, the participant received the task input (i.e. instructions, client profile, and hotel/room characteristics), and then they were informed of whether they would perform one or three tasks. There was no time limit to performing the task. This procedure was done three times with those participants who engaged in performing a sequence of three tasks, and once with those who performed one task. Performing the simple task took about 4.5 minutes, while the complex and +complex task each took about 7 minutes to perform. The overall pool of data consisted of 147 audio files. The data were recorded with several digital recorders Marantz PMD620.

8 Measures

Two fluency constructs were measured: speed fluency (unpruned speech rate, or 'Rate A', and pruned speech rate, or 'Rate B'; e.g. Robinson, 2007) and repair fluency (dysfluency ratio; e.g. Hsu, 2012). Both rates were calculated by dividing the total number of syllables by the total time, and multiplying it by 100. In the case of Rate A, the texts were unpruned, and they were pruned in the case of Rate B (all repetitions, restarts, and self-repairs were deleted from the narratives). Dysfluency ratio was calculated by dividing the total number of dysfluency markers (repetitions, false starts, and lexical and morphosyntactic repairs) by the total time, and multiplying it by 100.

Lexical complexity was measured via Guiraud's Index (e.g. Michel et al., 2007) and D index (e.g. Tavakoli, 2009) since these measures are not sensitive to text length (Read, 2000). Additionally, D calculates type–token ratio for different samples of words from a text instead of from the entire text.

Given that speech may get complexified at different levels (Norris & Ortega, 2009), three constructs of structural complexity were measured: overall complexity (mean length of AS-unit[4]), subordination (clauses[5]/AS-unit), and phrasal complexity (words/clause). AS-unit was chosen due to its suitability for spoken discourse. Clauses/AS-unit was chosen because it was anticipated that task design would prompt the use of subordination through having to engage in such mental operations as justifying, apologizing, and recommending. Words/clause is a unique measure insofar as it 'taps a more narrowly defined source of complexification' (Norris & Ortega, 2009, p. 561).

Concerning accuracy, three types of errors were calculated: lexical (word choice), morphosyntactic (syntax), and other errors (e.g. pragmatic). Two global accuracy measures were errors/AS-unit (e.g. Shiau & Adams, 2011) and errors/100 words (e.g. Sangarun, 2005), and one specific measure (target-like use of prepositions; TLU). Errors/AS-unit were calculated by dividing the total number of errors by the total number of AS-units. Errors/100 words were calculated by dividing the total number or errors by the total number of words, and multiplying it by 100. TLU was calculated by dividing the number of correctly supplied prepositions by the sum of the number of obligatory contexts and the number of incorrectly supplied prepositions, and multiplying it by 100. A handful of studies in the TBLT domain investigated a conceptually similar measure, target-like use of articles (e.g. Meraji, 2011).

9 Data transcription and coding

All data were transcribed in the CLAN mode of CHILDES and coded by the study author. In order to establish inter-coder reliability, 10% of the data were coded by a second coder (an experienced applied linguist). Cohen's kappa was calculated for each of the dependent variables: Rate A (0.91), Rate B (0.81), dysfluency ratio (0.78), errors/AS-unit (0.82), errors/100 words (0.82), target-like use of prepositions (0.88), mean length of clause (0.89), clauses/AS-unit (0.81), words/clause (0.89), Guiraud's Index (0.83), and D (0.97). The obtained values indicate strong inter-coder agreement, and are in line with values obtained in similar empirical research reporting analogical production measures (e.g. Kormos, 2011; Révész, 2011).

10 Data analysis

The independent variable in the current study was task sequencing, and the dependent variables were production dimensions: fluency (3 measures), accuracy (3 measures), and complexity (3 measures of structural complexity and 2 measures of lexical complexity). Two statistical techniques were employed: (1) descriptive statistics (means and standard deviations), and (2) inferential statistics to test for statistically significant differences. Concerning the latter, in order to determine whether the data followed normal distribution, skewness and kurtosis values were divided by the standard error of skewness or kurtosis, respectively. If the result was larger than 3.3, the data were considered not normally distributed. While some data followed the pattern of normal distribution (e.g. speech rate and lexical complexity), others did not (e.g. dysfluency ratio, accuracy, and structural complexity). Given this state of affairs, the use of non-parametric tests was warranted. Since the design of the study involved comparisons between groups of different speakers, Mann–Whitney U-test was employed to test for significant differences.

Alpha level in the study was set at .05 (Larson-Hall, 2000). The reported effect sizes are Cohen's d (Cohen, 1988) and they were calculated on the basis of means and standard deviations. Following Cohen (1988), the following benchmarks were employed when interpreting the magnitude of effect sizes: $d = .10$ to .29 or $d = -.10$ to $-.29$ (small effect size); $d = .30$ to .49 or $d = -.30$ to $-.49$ (medium effect size); $d = .50$ to 1.0 or $d = -.50$ to -1.0 (large effect size).

VI Results

This study set out to measure the performance on three tasks: a simple task, a complex one, and the most complex task, when they were performed in a simple–complex sequence versus in isolation. Table 3 presents the descriptive statistics (means and standard deviations) for the simple–complex sequencing group and for the individual task performance group, as well as inferential statistics (Mann–Whitney U-test). As can be observed in Table 3, significant differences in performance between the two groups were detected in three areas: fluency, accuracy, and structural complexity. The participants in the simple–complex sequencing group delivered speech faster than the individual task performance group on two of the three tasks (for +CT, $p = .019$ for Rate A, and $p = 0.19$ for Rate B; for ST, $p = .017$ for Rate A and $p = .023$ for Rate B). The obtained effect sizes range from 0.54 to 0.69, so they can be classified as medium to large, confirming a considerable difference to the advantage of the simple–complex sequencing group. However, the tendency is the opposite for dysfluency ratio: the speakers in the individual task performance condition produced significantly fewer dysfluencies than the

Table 3. Descriptive statistics (means and standard deviations) and inferential statistics (Mann–Whitney U-test, *p*-values and Cohen's *d*).

Dependent variable	Group	Simple task				Complex task				Most complex task			
		M	*SD*	*p*	*d*	*M*	*SD*	*p*	*d*	*M*	*SD*	*p*	*d*
Fluency:													
Rate A	IND	95.22	43.35	.017*	0.69	113.27	35.01	.460	0.3	98.38	34.99	.019*	0.65
	SCS	120.50	26.38			122.50	25.66			118.75		26.63	
Rate B	IND	89.46	42.39	.023*	0.61	102.98	32.88	.538	0.25	90.76	34.77	.019*	0.54
	SCS	111.22	25.88			110.35	24.33			107.26	24.49		
Dysfluency	IND	2.50	1.21	.001*	1.07	4.10	2.56	.000*	1.33	2.99	1.83	.007*	0.79
	SCS	4.23	1.94			8.41	3.78			4.50	1.96		
Accuracy:													
Errors/AS-unit	IND	1.60	1.17	.011*	−0.72	1.14	0.56	.121	−0.28	1.10	0.57	.148	−0.31
	SCS	0.97	0.49			0.96	0.71			0.91	0.62		
Errors/100 words	IND	17.32	10.17	.010*	−0.79	12.94	6.84	.094	−0.44	12.56	7.05	.109	−0.43
	SCS	10.62	6.17			10.17	5.52			9.66	6.16		
Target-like use of	IND	51.45	22.84	.008*	0.79	55.72	13.59	.001*	1.03	69.31	18.18	.104	0.44
prepositions	SCS	68.12	18.70			71.75	17.16			76.39	13.15		
Complexity: Structural:													
Words/AS-unit	IND	9.14	1.36	.924	−0.09	8.98	1.27	.652	0.04	9.14	1.37	.656	−0.09
	SCS	9.02	1.18			9.05	1.89			9.02	1.23		
Words/Clause	IND	7.68	1.40	.647	−0.33	7.17	1.98	.460	−0.23	6.57	0.58	.047*	0.75
	SCS	7.22	1.36			6.77	1.42			7.60	1.84		
Clause/AS-unit	IND	1.21	0.18	.536	0.26	1.31	0.28	.918	0.19	1.40	0.22	.022*	−0.73
	SCS	1.34	0.66			1.36	0.24			1.23	0.24		
Complexity: Lexical:													
Guiraud's Index	IND	5.06	0.83	.246	0.40	5.22	0.47	.330	−0.24	5.33	0.74	.440	−0.31
	SCS	5.36	0.63			5.08	0.67			5.12	0.59		
D	IND	39.31	10.81	.268	0.45	46.43	11.06	.943	0.05	49.44	11.87	.357	0.23
	SCS	44.32	11.30			47.07	12.10			52.32	12.31		

Notes. IND = individual task performance group, SCS = simple–complex sequencing group. Rate A = unpruned speech rate; Rate B = pruned speech rate.* α significant at $p < .05$.

simple–complex sequencing group (p = .007 for +CT, p = .000 for CT, and p = .001 for ST). The obtained effect sizes are very large, with a difference of over one standard deviation between the two groups in the simple and the complex task, to an exponential advantage to the learners in the baseline condition.

Regarding accuracy, the error rate in the simple–complex condition group was significantly lower than in the individual task performance group on the simple task (p = .011 for errors/AS-unit, p = .010 for errors/100 words, and p = .008 for TLU), and on the complex task in the specific measure (p = .001 for TLU). Considerably more target-like production in the case of the simple–complex sequencing group is further evidenced in large effect sizes obtained for all statistically significant results. Although no statistically significant results were detected in the case of the most complex task, the speakers in the simple–complex sequencing group delivered more target-like production in this task as well.

Concerning structural complexity, significant differences were detected only in the case of the most complex task, with the results pointing to advantages in both groups depending on the measure. While clause-internal expansion was greater in the case of the simple–complex sequencing group than in the individual task performance group (p = .047 for words/clause), subordination index was higher in the case of the latter (p = .022 for clauses/AS-unit). Both results were accompanied by large effect sizes, further corroborating a considerable difference in performance between the two groups. In the remainder of the measures and tasks, the results were mixed and highly unsystematic.

Finally, no significant differences were detected for lexical complexity. As can be seen in descriptive statistics, the *D* index showed greater lexical complexity in the simple–complex group on all tasks, while Guiraud's index detected this group's advantage compared to the individual task performance group only in the simple task.

To sum up the results, the speech of the participants in the simple–complex sequencing condition was characterized by greater speed (all tasks), lower error rate (simple task and specific measure in complex task), and greater structural complexity at clausal level in the most complex task. By contrast, the participants in the individual task performance group produced fewer dysfluencies across all tasks, and their speech was structurally more complex at the subordination level in the most complex task.

VII Discussion

The role of sequencing in production in this study was investigated by having the same speaker perform a sequence of three tasks of differing cognitive complexity levels on the one hand, and on the other hand by having each of these three tasks performed by three different groups of speakers. Major findings from this experiment can be summarized as follows:

- Sequencing tasks from simple to complex led to a higher speech rate (ST and +CT), partially enhanced accuracy (all measures in ST and TLU in CT), and to greater structural complexity in terms of clause-internal expansion (words/clause in +CT).
- Individual task performance led to a considerably lower dysfluency ratio (all tasks), and to greater structural complexity in terms of subordination ratio (clauses/AS-unit in +CT).

- The two conditions exhibited different patterns in structural complexity.
- No notable differences emerged in lexical complexity.

In terms of speech rate, from a speech production perspective the participants in the simple–complex sequencing group perhaps produced their output faster because each subsequent task required systematically less engagement in the process of message conceptualization, given that the sequence of three tasks followed the same procedure, instructions, and analogous task input. After performing ST, they were faced with a relatively similar task input in the two other tasks. Therefore, the procedures encountered in the first task were likely transferred to the performance of CT and +CT. Perhaps this transfer of knowledge to a new context eliminated the need to conceptualize the message, and required less planning at micro and macro levels. It is therefore possible that the available resources were channeled towards more automatized message delivery. From the perspective of instance theory (Logan, 1988), the difference in speech rate between the two groups could be attributed to memory. As the participants engaged in more input reception, and more production of output, the formulas they had had to use became increasingly readily available. Rather than having to be selected from the mental lexicon, they were retrieved from memory in the form of previously employed chunks, alleviating the mental burden associated with the search process. It should be acknowledged that although the tasks were manipulated in terms of their cognitive complexity, inevitably in the simple–complex sequencing group to some extent the participants engaged in task repetition. The participants in the individual task performance condition received substantially less (a third, to be exact) exposure to output than their counterparts in the simple–complex condition. At least some studies in the area of task repetition point to the fact that repeating tasks – either in terms of their content or procedure – supports oral development in the area of fluency (e.g. Bygate, 2001; Bygate & Samuda, 2005; Lynch & Maclean, 2000, 2001). It could therefore be the case that the variable responsible for higher speed rate in the simple–complex sequencing group could be simple–complex sequencing, task repetition, or a combination thereof.

These insights notwithstanding, it remains intriguing why the participants in the simple–complex sequencing group delivered their speech faster also in the ST. One explanation is that these participants had only 1 minute between the tasks, which may have oriented them to complete each task fast to prevent fatigue. The participants in the individual task performance group had only one task to complete so they may have felt less time pressure and they were more likely to take their time. In both groups, the participants were informed about how many tasks they would perform right before engaging in the first task.

Fast speech delivery in the simple–complex condition occurred concurrently with considerably more instances of dysfluencies than in the individual task performance condition. In fact, the pattern observed across tasks was that the participants in the simple–complex sequencing group produced roughly double the number of dysfluencies than those in the individual task performance group. Given that across tasks the participants in the simple–complex sequencing condition delivered their speech faster than those in the individual task sequencing group, dysfluency behaviour might be related to, or dependent on, speech rate: the faster the rate at which output is delivered, the greater the occurrence of dysfluency phenomena. In other words, there may have been a trade-off between the speed of delivery and repair fluency.

Another explanation for this finding lies in the way dysfluency was operationalized in this study. Some researchers suggested that dysfluency, and in particular self-repairs, should be conceived of as a measure of accuracy, as evidence of focus on form, rather than as a fluency measure (e.g. Gilabert 2007; Lyster & Ranta, 1997; Swain, 1998). In the data analysis in the current study, self-repairs were not disentangled from other dysfluency phenomena (repetitions and false starts). If self-repairs had been operationalized as an accuracy index, the obtained results could potentially be interpreted as simple–complex sequencing leading to more accurate linguistic behaviour, rather than more dysfluent behaviour. Since the results of the employed accuracy measures lend partial support to the idea that simple–complex sequencing leads to greater accuracy (such was the case on all measures in the simple task and on TLU in the complex task), the interpretation of dysfluency as an accuracy measure would add evidence to the role of simple-task sequencing in promoting accuracy. Regarding accuracy, a noteworthy pattern in findings lies in the fact that the participants in all groups exhibited a more accurate behaviour on each subsequent task, on all measures (i.e. participants in both groups displayed a lower error rate on +CT than on CT, and still lower on CT than on ST; the same pattern holds true for TLU). Therefore, improvements in accuracy were the function of increases in task demands. This pattern of findings suggests that perhaps the agent responsible for more target-like behaviour was the cognitive complexity of tasks, and not the sequencing order in which the tasks were performed.

Regarding syntactic complexity, while the participants in the individual task performance condition complexified their speech at the level of subordination and decomplexified it at the level of words/clause, the opposite pattern could be observed in the simple–complex sequencing group. These speakers produced longer clauses, with a simultaneous decrease in clauses/AS-units. This performance pattern is surprising insofar as complexifying at the level of clause, rather than at the subordination level, is characteristic of greater proficiency in the L2: as interlanguage develops, in terms of syntactic complexity learners may make a shift from complexifying at the subordination level and move on to a higher-level complexification, that is, at the clausal level. However, in the current study complexification at the clausal level probably emerged as a result of the interaction of two factors: the speakers' proficiency level, and increasingly higher cognitive demands. Performing a sequence of tasks therefore led to a deeper-level, more advanced structural complexification of speech, typically associated with different levels of L2 competence.

In terms of the SSARC model of pedagogic task sequencing, in the simple–complex sequencing group, ST and CT may have served the function of scaffolding. In these tasks, the speakers perhaps relied on old knowledge and familiar structures as a way of consolidating their interlanguage state. This pattern matches what Robinson (2010) described as the 'simple, stable <attractor> state of current interlanguage', associated with cognitively simple tasks. It was in response to the most complex task that the speakers stretched their syntactic repertoire beyond the simple subordination and moved on to deeper levels of structural complexification. Again, in Robinson's terms, this means that the most complex task promoted 'restructuring of the current interlanguage system … introducing maximum complexity' (Robinson, 2010, p. 247). However, the three tasks had the opposite effect on those speakers who performed only one task: these participants complexified their speech at the level of subordination, and simultaneously decomplexified it at the clausal level. When these participants were faced with a more cognitively demanding task, they produced shorter clauses and more subordination.

The experiment reported in the current study aimed to put to test sequencing tasks from simple to complex, as postulated by the SSARC model of pedagogic task sequencing. The findings obtained in this study indicated that such sequencing triggered a higher speech rate (ST and +CT), greater accuracy (all measures in ST and TLU in CT), greater structural complexity in terms of clause-internal expansion (words/clause in +CT), and greater dysfluency. The findings obtained, which compared three tasks performed by the same speaker and performed by different speakers, revealed that the manipulation of a task's cognitive load certainly has a different effect on language production when tasks are performed in a sequence versus in the absence of any other task.

VIII Conclusions

As the results of this study demonstrated, simple–complex sequencing, and associated with it greater amount of exposure, promote greater fluency, accuracy, and higher-order syntactic restructuring. The implication is that, in the classroom context, task-based lessons should include such sequences of tasks in order to promote the development of these aspects of performance, given that tasks designed and implemented in this way provide a more effective way of targeting these production dimensions than performing the same tasks in an individual manner. On the other hand, having learners carry out one-off tasks can be a sound pedagogic choice when it comes to producing oral speech with fewer reformulations, restarts, and repetitions.

In this study, learners in the simple–complex sequencing condition performed the tasks at 1-minute intervals. Future research should verify how such massed, intense exposure compares to alternative 'spaced' intervals in terms of effects on oral performance, i.e. leaving learners more time between the tasks. Obtaining more insights into this issue through research has the potential of further informing teachers' pedagogical choices and contributing to the debate about optimal sequencing decisions.

As already stated in Section VII, it must be kept in mind that in the current study there were task repetition effects in one group, but not in the other. In order to disentangle the two types of effects – task complexity and task repetition - a design comparing simple–complex sequence with alternative sequencing orders is warranted in order to corroborate the findings obtained. The unexplored issues mentioned will hopefully be addressed by future empirical explorations, advancing our understanding of the optimal task sequencing order or orders, and bringing us closer to a more robust conceptualization of task sequencing.

Acknowledgements

The data presented in this paper comes from my doctoral dissertation. I would like to thank my advisors, Dr. Roger Gilabert and Dr. John Norris, for their ideas and feedback on this study. I would also like to thank two anonymous reviewers for their insightful comments on the first draft of this paper.

Conflict of Interest

Also, if this study is part of a larger study or if you have used the same data in whole or in part in other papers, both already published or under review please state where the paper is published and describe clearly and in as much detail as you think necessary where the similarities and differences are and how the current manuscript makes a different and distinct contribution to the field.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. e = mental effort; s = simple task demands; c = complex task demands; rdisp = resource dispersing dimensions of tasks; rdir = resource directing dimensions of tasks; ^{n} = potential number of practice opportunities on tasks.
2. For a detailed operationalization of cognitive complexity in the three tasks, the instructions and task input, see Malicka, Gilabert, and Norris (2017).
3. The procedure in the pilot study was the same as that employed in the main experiment described in Section V.7.
4. The definition of AS-unit was adopted from Foster et al. (2000, p. 365): 'An AS-unit is a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either … An independent clause will be minimally a clause including a finite verb.'
5. The definition of 'clause' was adopted from Foster et al. (2000, p. 365), and it was defined as consisting 'minimally of a finite or non-finite Verb element plus at least one other clause element (Subject, Object, Complement, or Adverbial)'.

References

Ahmadian, M. (2012). The effects of guided careful online planning on complexity, accuracy and fluency in intermediate EFL learners' oral production: The case of English articles. *Language Teaching Research*, *16*, 129–149.

Baralt, M. (2010). Task complexity, the Cognition Hypothesis and interaction in CMC and FTF environments. Unpublished PhD dissertation, Georgetown University, Washington, DC, USA.

Baralt, M. (2013). The impact of cognitive complexity on feedback efficacy during online versus face-to-face interactive tasks. *Studies in Second Language Acquisition*, *35*, 689–725.

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

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

Benson, S.D. (2015). Task-based language teaching: An empirical study of task transfer. *Language Teaching Research*, *20*, 341–365.

Brindley, G. (1987). Factors affecting task difficulty. In Nunan, D. (Ed.), *Guidelines for the development of curriculum Resources* (pp. 45–56). Adelaide: Adelaide National Curriculum Resource Centre.

Brown, G.K., Malmkjaer, K., & Williams, J. (Eds.). (1984). *Teaching talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In Bygate, M., Skehan, P., & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Harlow: Pearson Longman.

Bygate, M., & Samuda, V. (2005). Integrative planning through the use of task-repetition. In Ellis, R. (Ed.), *Planning and task performance in a second language: Volume 11* (pp. 37–74). Amsterdam: Benjamins.

Candlin, C. (1987). Towards task-based language learning. In Candlin, C., & D. Murphy (Eds.), *Language learning tasks* (pp. 5–22). Englewood Cliffs, NJ: Prentice-Hall.

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

Craik, F., & Lockhart, R. (1972). Level of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*, 354–375.

Gilabert, R. (2007). Effects of manipulating task complexity on self-repairs during L2 oral production. *IRAL: International Review of Applied Linguistics in Language Teaching*, *45*, 215–240.

Hsu, H. (2012). Investigating the effects of planning on L2 text chat performance. *CALICO Journal*, *29*, 619–638.

Hsu, H.C. (2015). The effect of task planning on L2 performance and L2 development in text-based synchronous computer-mediated communication. *Applied Linguistics*, *38*, 359–385.

Hulstijn, J. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of rehearsal, elaboration and automaticity. In Robinson, P. (Ed.), *Cognition and second language instruction* (pp. 349–381). Cambridge: Cambridge University Press.

Hulstijn, J. (2003). Incidental and intentional learning. In Doughty, C., & M.H., Long (Eds.), *Handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.

Ishikawa, T. (2011). Examining the influence of intentional reasoning demands on learner perceptions of task difficulty and L2 monologic speech. In Robinson, P. (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 307–329). Amsterdam: John Benjamins.

Jackson, D.O., & Suethanapornkul, S. (2013). The Cognition Hypothesis: A synthesis and meta-analysis of research on second language task complexity. *Language Learning*, *63*, 330–367.

Kim, Y., & Payant, C. (2014). A pedagogical proposal for task sequencing: An exploration of task repetition and task complexity on learning opportunities. In Baralt, M., Gilabert, R., & P.,

Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 151–177). London: Bloomsbury.

Kim, Y., & Taguchi, N. (2015). Promoting task-based pragmatics instruction in EFL classroom contexts: The role of task complexity. *The Modern Language Journal*, *99*, 656–677.

Kim, Y., & Tracy-Ventura, N. (2011). Task complexity, language anxiety, and the development of the simple past. In Robinson, P. (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 287–306). Amsterdam: John Benjamins.

Kim, Y., Payant, C., & Pearson, P. (2015). The intersection of task-based interaction, task complexity, and working memory: L2 question development through recasts in a laboratory setting. *Studies in Second Language Acquisition*, *37*, 549–581.

Kormos, J. (2011). Task complexity and linguistic and discourse features of narrative writing performance. *Journal of Second Language Writing*, *20*, 148–161.

Kormos, J., & Trebits, A. (2011). Working memory capacity and narrative task performance. In Robinson, P. (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 267–287). Amsterdam: John Benjamins.

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

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

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. London: Routledge.

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

Logan, G.D. (1988). Toward an instance theory of automatisation. *Psychological Review*, *95*, 492–527.

Lynch, T., & Maclean, J. (2000). Exploring the benefits of task repetition and recycling in classroom language learning. *Language Teaching Research*, *4*, 221–250.

Lynch, T., & Maclean, J. (2001). 'A case of exercising': Effects of immediate task repetition on learners' performance. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 141–162). Harlow: Pearson Longman.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, *19*, 37–66.

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 Academic.

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

Malicka, A., Gilabert, R., & Norris, J. (2017). From needs analysis to research tasks: Insights from an English for specific purposes context. *Language Teaching Research*, *21*, 1–29.

Meraji, S.R. (2011). Planning time, strategy use, and written task production in a pedagogic vs. a testing context. *Journal of Language Teaching and Research*, *2*, 338–352.

Michel, M. (2011). Effects of task complexity and interaction on L2 performance. In Robinson, P. (Ed.), *Second language task complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 141–173). Amsterdam: John Benjamins.

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

Murphy, L., & Roca de Larios, J. (2010). Searching for words: One strategic use of the mother tongue by advanced Spanish EFL writers. *Journal of Second Language Writing*, *19*, 61–81.

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.

Nuevo, A.-M., Adams, R., & Ross-Feldman, L. (2011). Task complexity, modified output, and L2 development in learner–learner interaction. In Robinson, P. (Ed.), *Second language task* *complexity: Researching the Cognition Hypothesis of language learning and performance* (pp. 175–201). Amsterdam: John Benjamins.

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

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

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

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, *95*, 162–181.

Révész, A., Michel, M.C., & Gilabert, R. (2012). Measuring cognitive task demands: A step forward in task complexity research. Unpublished paper presented at the 22nd European Association of Applied Linguistics (EuroSLA) Conference, Poznan, Poland.

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

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

Robinson, P. (2003). Attention and memory during SLA. In Doughty, C., & Long, M.H. (Eds.), *The handbook of second language acquisition* (pp. 631–678). Oxford: Blackwell.

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

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *IRAL: International Review of Applied Linguistics in Language Teaching*, *45*, 193–213.

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

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In Ellis, R. (Ed.), *Planning and task performance in a second language* (pp. 111–141). Amsterdam: John Benjamins.

Sasayama, S., Malicka, A., & Norris, J. (2015). Primary challenges in cognitive task complexity research: Results of a comprehensive research synthesis. Unpublished paper presented at the 6th Biennial International Conference on Task-Based Language Teaching (TBLT), Leuven, Belgium.

Schank, R. (1999). *Dynamic memory revisited*. New York: Cambridge University Press.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals and understanding*. Hillsdale, NJ: Erlbaum.

Shiau, Y.S., & Adams, R. (2011). The effects of increasing reasoning demands on accuracy and complexity in L2 oral production. *University of Sydney Papers in TESOL*, *6*, 121–146.

Skehan, P. (1996). A framework for the implementation of task based instruction. *Applied Linguistics*, *17*, 38–63.

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

Swain, M. (1998). Focus on form through conscious reflection. In Doughty, C., & J., Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–81). New York: Cambridge University Press.

Tavakoli, P. (2009). Assessing L2 task performance: Understanding effects of task design. *System: An International Journal of Educational Technology and Applied Linguistics*, *37*, 482–495.

Tavakoli, P., & Foster, P. (2008). Task design and second language performance: The effect of narrative type on learner output. *Language Learning*, *58*, 439–473.
Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomic Science*, *3*, 159–177.
Wickens, C. D. (2007). Attention to the second language. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*, 177–191.

Appendix 1.

Time judgment task and affective variables questionnaire.

Time judgment task

How long did it take you to complete the task?

Estimated time
__________ in minutes and _________ seconds

Real time
____

Affective variables questionnaire

Please tick the appropriate box on the scales below.

1. This task required …

   … no mental effort at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ extremely great mental effort

2. This task was …

   … not difficult at all ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ extremely difficult