JESLA Journal of the European Second Language Association

Santos, S. (2018). Effects of task complexity on the oral production of Chinese learners of Portuguese as a foreign language. *Journal of the European Second Language Association,* 2(1), 49–62, DOI: https://doi.org/10.22599/jesla.40

RESEARCH

Effects of task complexity on the oral production of Chinese learners of Portuguese as a foreign language

Sara Santos

This study aims to investigate the effects of task complexity on the oral performance of Chinese learners of European Portuguese as a foreign language (PFL). In the current investigation, the cognitive task demands were manipulated along the resource-directing and resource-dispersing dimensions of the Triadic Componential Framework (Robinson, 2001, 2003, 2005, 2010, 2011, 2015). The variables chosen were respectively ±few elements and ±planning time. Using a 2 × 2 design, 39 university learners of PFL performed two monologic information-giving oral tasks: a simple task (two elements) and a complex task (six elements). Half of the learners (n = 19) were given pre-task planning time, but for the other half (n = 20) the planning time was removed. The order of the tasks was counterbalanced. The participants' output was analysed by general and specific measures of syntactic complexity and accuracy, lexical diversity and fluency (CALF). Pre-task planning time had significant effects on accuracy. Increasing the number of elements of the task led to greater accuracy and lexical diversity and longer clause length. There were no significant effects on fluency. These findings partially support Robinson's Cognition Hypothesis. To the best of our knowledge, this study is the first CALF study to measure the oral performance of learners of PFL. These results provide new insights for research and learning in the field of instructed second language acquisition (ISLA).

Keywords: task complexity; oral performance; PFL learners

1. Introduction

1.1. Cognitive task complexity and attentional resources

In the last few decades, the use of tasks in L2 classrooms has been widely proposed as an optimal tool to promote language development and acquisition. Several studies have investigated the impact of the cognitive complexity of a task on the way learners allocate their attention towards language while performing the task. In the literature, there are two competing models of attention allocation: the Limited Attention Capacity (LAC) approach (Skehan, 1998, 2009, 2014, 2015) and the Cognition Hypothesis (CH) (Robinson, 2001, 2003, 2005, 2011, 2015). The former framework claims that different performance areas (complexity, accuracy, lexis and fluency) might compete for resources when a task's cognitive complexity increases, due to limited attentional capacity. According to the LAC approach, if production is linguistically more complex, accuracy and fluency may not also be elevated; thus, increased complexity 'might be associated with lower fluency, or raised accuracy with lower complexity' (Skehan, 2015, p. 125). Following this perspective, fluency and complexity often go together, as do accuracy and fluency, but the least likely association is elevated complexity and accuracy (Skehan, 2009, 2014, 2015). The purpose of LAC research is to explore how task characteristics and conditions can mitigate avoidable trade-off effects between complexity and accuracy, as these performance areas tend to compete for attentional resources (Wang & Skehan, 2014). In contrast, the CH proposes a multiple resources attentional model and claims that both linguistic complexity and accuracy increase during task performance. Within the CH, Robinson (2010, 2015) presented a taxonomy of task characteristics—the Triadic Componential Framework (TCF)—and the stabilize, simplify, automatize, restructure and complexity (SSARC) model for pedagogic task sequencing, which is based on the premise that learners should perform simple tasks on relevant parameters first, and then the cognitive demands of the task should be increased on subsequent versions for cumulative learning. The TCF classifies task characteristics according to three different factors: cognitive, interactive and learner factors. By making use of cognitive factors, task complexity can be manipulated by syllabus task designers to push the learner's output. In his model, Robinson distinguished between resource-directing and resource-dispersing factors. According to his framework, increasing the complexity of a task along the resource-directing dimensions (±here and now; ±few elements;

Department of Portuguese, Faculty of Arts and Humanities, University of Macau, CN

saras@umac.mo

±spatial reasoning; ±causal reasoning; ±intentional reasoning; ±perspective taking) will direct the attention of the learner to the form of the target language and consequently will lead to a more complex and accurate output. In contrast, increasing the demands of a task along resource-dispersing variables (±planning time; ±single task; ±task structure; ±few steps; ±independency of steps; ±prior knowledge) will disperse the attentional and memory resources of the L2 learner with negative consequences for production, in respect to all components (linguistic complexity, accuracy and fluency). Robinson's (2010, 2015) SSARC pedagogic model for task design and sequencing is based on the premise that the cognitive complexity of the task is to be increased first along the resource-dispersing dimensions and only afterwards along the resource-directing dimensions.

This study aims to analyse the effects of manipulating the cognitive complexity of a task along the resource-directing factor [±few elements] and along the resource-dispersing variable [±planning time] on the oral production of Chinese learners of European PFL, an under-researched population in CALF literature. In the next section, a brief review of previous studies investigating the factor ±planning time and studies investigating the variable ±few elements is presented.

2. Previous research

2.1. Studies of ± planning time

Several studies have investigated the effects of pre-task planning time on learners' oral performance. Presenting a descriptive synthesis of planning time studies, Ellis (2009) outlined some conclusions: (i) overall, strategic planning has positive effects on fluency; (ii) in respect to accuracy and complexity the effects of planning are more variable (although clearer for complexity); (iii) strategic planning has a stronger impact on fluency and complexity, suggesting that learners prioritise what they want to say, that is, the conceptualisation, rather than the formulation of a specific linguistic plan.

Skehan and Foster (1997), Mehnert (1998), Ortega (1999) and Wang (2014) confirmed that pre-task planning time had positive effects on fluency and syntactic complexity. Although in the first two studies the hypothesis of a competition for attentional resources between accuracy and complexity was supported, Ortega's (1999) accuracy results were mixed. Yuan and Ellis (2003) found that strategic planning positively affected syntactic complexity, and Guará-Tavares (2008, 2011) reported gains in respect to accuracy. Along the same lines, Mochizuki and Ortega (2008) showed that under guided planning conditions, learners' output was more accurate. Gilabert (2005) found that planners produced more fluent and lexically diverse speech. It should be noted that the participants in all these studies were learners of EFL, except for Mehnert (1998) and Ortega (1999), who investigated the oral performance of learners of German and Spanish, respectively. The informants were mostly from different L1 backgrounds, but Wang (2014) and Yuan and Ellis (2003) studied the oral L2 performance of Chinese native speakers. In most of these experiments, the learners were given ten minutes for pre-task planning time, but Mochizuki and Ortega (2008) gave only five minutes, and in Mehnert (1998) three experimental groups were given one, five and ten minutes, respectively. Questioning Robinson's distinction between resource-directing and resource-dispersing characteristics, Skehan (2009, 2014, 2015) claimed that recent LAC research had shown that two resource-dispersing features, i.e., planning time and task structure, had been mis-analysed, as their impact could result in a joint increase of complexity and accuracy, as Tavakoli and Skehan's (2005) study demonstrated. Skehan (2015) considered that planning was not a monolithic category (as sometimes it could meet the criteria of a resource-directing variable). This controversy and the lack of clear supportive evidence for the CH predictions justifies this study of planning, as some clarification is needed. Although previous planning studies yielded consistent results, some ambiguities remain concerning the actual impact of this task complexity feature on L2 learners' oral production.

2.2. Studies of ±number of elements

Citing Sasayama, Malicka and Norris's (2013, 2014, 2015) unpublished research synthesis, Sasayama (2015) found that the ±few elements factor was the most common operationalisation within Robinson's resource-directing variables. However, the studies included in this synthesis yielded inconsistent findings on the impact of raising the elements of a task on CALF measures (namely in respect to the first three dimensions), because according to Sasayama (2015) the increased task complexity results varied: (i) only some studies found that accuracy was increased; (ii) syntactic complexity was positive, null or negative across studies; and (iii) two studies showed a joint increase in accuracy and lexical diversity. The impact on fluency has been more consistent, as with a greater number of elements fluency decreased in all studies. Bearing these differences in mind, this section reviews studies in which the variable ± elements was manipulated as one of the independent variables.

Robinson (2001), Michel, Kuiken and Vedder (2007, 2012), Michel (2011), Révész (2011) and Oh and Lee (2012) investigated the effects of task complexity (±few elements) and interaction (±monologic). Robinson (2001) measured the output of 44 Japanese L1 students while performing two city map tasks. The results confirmed effects on lexical variety and on fluency. Michel et al. (2007, 2012) and Michel (2011) increased the cognitive demands of two argumentative tasks. In Michel et al. (2007), the participants (Turkish and Moroccan learners of Dutch) were given a full-colour leaflet with two (simple task) or six (complex task) electronic devices. Learners were more accurate and less fluent on complex tasks, but there were no significant main effects on syntactic complexity. Michel (2011) reported different findings: increasing the number of elements affected lexical diversity, but accuracy and fluency were not affected. Révész (2011) and Oh and Lee (2012) increased the cognitive complexity of a task along the number of elements in different ways: the former used an argumentative task–the simple and complex version diverged with respect to the amount of economic resources and the number of projects to be allocated (three vs. six)–and the latter chose a narrative task, with more

complex storylines and characters in the complex version of the task. Forty-three learners of English, with different L1 backgrounds, participated in Révész's (2011) classroom-based study. Regarding general production measures, their speech was lexically more diverse and accurate, but syntactically less complex. The specific measure used showed that participants were more likely to use complex conjoined clauses while performing the complex task. Oh and Lee (2012) examined the oral production of 40 Korean university learners of English. There were no overall significant effects on linguistic complexity, accuracy or fluency. Kuiken and Vedder (2011, 2012) studied the effects of increasing the number of elements and proficiency level in a written vs. oral argumentative task performed by 44 Dutch students of Italian L2. The results contradicted the CH, as in the oral mode the manipulation of the factor ±elements led to gains in accuracy but less syntactic complexity. Sasayama (2015) researched the effects of task complexity along ±few elements; Malicka (2014) did likewise and also investigated the impact of ±spatial reasoning dimensions. Instead of using a dichotomous approach, these two studies used tasks with multiple levels of cognitive complexity: the former had four tasks and the latter three. Sasayama (2015) found that the most complex task (with more elements: nine characters in a narrative) did not elicit the best L2 performance (in terms of accuracy and linguistic complexity); the best results were elicited by the second-simplest task according to measurements (although it was designed to be the third-simplest). In contrast to Robinson's predictions, Sasayama (2015, 2016) found that the resource-directing feature (number of elements) could have deleterious effects. The findings showed that the four tasks posed a mix of both extraneous and facilitative load, though to clearly different degrees. Malicka (2014) found that increasing task complexity along the number of elements promoted more accurate and lexically diverse production, so the CH was only partially confirmed.

Levkina (2008), Levkina and Gilabert (2012) and Sasayama and Izumi (2012) manipulated task complexity along the resource-directing ±few elements variable and the resource-dispersing ± planning time factor, based on Robinson's CH, as in this study. The participants were all learners of EFL: in Levkina (2008), they had different L1 backgrounds; in Levkina and Gilabert (2012), the learners were Spanish and Russian L1 speakers; and Sasayama and Izumi (2012) investigated the performance of Japanese L1 high school students. Levkina (2008) and Levkina and Gilabert (2012) used a Latin Square design, and the participants were asked to perform four tasks under four different conditions. Participants in both studies received four full-colour leaflets with two (simple task) or six (complex task) holiday destinations or apartment descriptions, but Sasayama and Izumi (2012) manipulated two monologic narrative tasks. In the three experiments, participants had five minutes for the planned condition. According to Sasayama and Izumi (2012), planners' performance was less fluent. Planning had positive results: on lexical diversity and syntactic complexity in Levkina (2008), on lexical diversity and fluency in Levkina and Gilabert (2012) and on syntactic complexity in Sasayama and Izumi (2012). Regarding the manipulation of the factor '±few elements', the three studies confirmed a decrease in fluency when learners performed a task with more elements. Levkina (2008) also showed a negative impact on syntactic complexity and Sasayama and Izumi (2012) found a negative impact on accuracy. The latter study reported that a greater number of elements in a task positively affected the specific measure of syntactic complexity, and Levkina and Gilabert (2012) showed positive effects on lexical diversity. Levkina (2008) found combined effects of '±planning time' and '±few elements' on lexical diversity and Levkina and Gilabert (2012) reported significant overall combined effects for not only lexical diversity but also fluency. As is demonstrated by this brief literature review, research findings are somewhat mixed and not conclusive; neither Robinson's nor Skehan's framework has been confirmed or disconfirmed. Note that many researchers have chosen an argumentative task type, although the operationalisation of the variable ±few elements has varied.

3. The present study

As no study has investigated the effects of task complexity on the oral performance of Chinese learners of PFL within the CH, the study reported here investigated the impact of manipulating the two independent variables ±few elements and ±planning time. The research questions (RQ) are as follows.

3.1. Research questions and hypotheses

RQ1: What are the effects of manipulating the cognitive task complexity along the resource-dispersing factor (±planning time) on the oral production of Chinese learners of PFL?

RQ2: What are the effects of increasing the cognitive task complexity along the resource-directing variable (±few elements)?

RQ3: To what extent does the simultaneous manipulation of planning time and number of elements affect the oral performance of these learners?

Based on the claims of the CH, the following hypotheses (H) were formulated for each of the research questions.

H1: Increasing the cognitive task complexity along the ±planning time factor will have negative effects on all dimensions of L2 oral production.

H2: Increasing the number of elements of a task will result in a less fluent but more accurate, complex and lexically diverse speech.

H3: The attention-directing effect of performing a complex task along the number of elements will be boosted by decreased complexity along the resource-dispersing dimension (+planning).

4. Methodology

4.1. Participants

Thirty-nine Chinese university learners of PFL participated voluntarily in the study. Twenty-three students were Cantonese native speakers, twelve students were

Mandarin native speakers and four students were bilingual (Cantonese and Mandarin). Their mean age was 20.59; 71.8% were female and 28.2% were male. They were undergraduate students majoring in Portuguese. Concerning Portuguese language learning, they had the same educational background: 840 hours of formal language learning in PFL. The level of proficiency was between A2 and B1, as the results of the standardised exam *DEPLE* for PFL for the B1 CEFR level of proficiency ranged between 44% and 74%. Participants were evenly assigned to the experimental conditions.

4.2. Tasks and procedures

Two sets of monologic information-giving tasks (appendix 1) were designed on the same topic: travelling and holidays. Learners received a full-colour leaflet with two holiday destinations (countries or cities) for the simple task or six for the complex task. The input was mainly visual, to avoid lexical support in L2. The simple version of the task included the name of each country/city, three pictures, the price, the airline company and the number of travelling days. The complex task offered six destinations, four of which were in Asia, as students were more familiar with Asian cultures. Each destination had five pictures, the number of travelling days and dates, the price and four hotel icons. After examining the visual input, learners were asked to leave a message in a friend's WeChat giving information about all possible holiday destinations. All participants performed one simple and one complex task. The order of the tasks was counterbalanced to avoid possible carryover effects from one experimental condition to another. Half of the learners (n = 19) had five minutes to plan the task, and the other half (n = 20) were only given 30 seconds to look briefly at the leaflet. Learners performed the tasks in a language laboratory environment. To validate the construct of task complexity, students were asked to rate their performance on a seven-point Likert scale affective variables questionnaire (AVQ) after completing each task.

4.3. Dependent variables – CAF measures

Bearing in mind that CAF measures are multidimensional constructs (Housen, Kuiken & Vedder, 2012; Michel, 2017; Norris & Ortega, 2009), both general and specific measures were chosen to quantify learners' oral production. Linguistic complexity combined five measures: syntactic complexity was measured by words per clause (clause length), clauses per AS-unit and coordinate clauses per AS-unit; lexical diversity was measured by Guiraud's Index and VOCD, computed by CLAN. For accuracy, two general measures (percentage of error-free clauses per total clauses and errors per 100 words) and four specific measures were chosen (lexical errors per AS-unit, omissions per AS-unit, morphosyntactic errors per AS-unit and the percentage of self-repairs per total errors). For fluency, three measures were calculated: two measures of speech rate (rate A, i.e., the ratio of words per minute in unpruned speech, and rate B, i.e., the ratio of words per minute in pruned speech) and one measure of fluency repair (number of repetitions, self-repairs, reformulations and false starts per minute).

4.4. Transcription and coding

To transcribe and code the data, the CLAN program (MacWhinney, 2000) was used. The selected basic unit for analysis was the AS-unit (Foster, Tonkyn & Wigglesworth, 2000). The transcription of the speech samples was carried out by the researcher and a research assistant, who were native speakers of European Portuguese. The researcher checked all of the transcripts and coding. Interrater reliability was assessed by means of percentage agreement on 10% of the data (randomly selected), and reached 97.7%.

4.5. Statistical analyses

The statistical analyses were carried out with SPSS 24.0 for Windows. To address the research questions, a repeated measures analysis of variance (ANOVA) was performed with ±task complexity as the within-subjects factor and ±planning time as the between-subjects variable. The results reported here should be interpreted with caution as multiple ANOVAs were computed in this study. The alpha-level was set at $p < 0.05$. Partial eta square (η_p^2) effect sizes are reported for reference, as they were computed by the SPSS; Cohen's d values were also calculated, as they are commonly used in SLA research and can help avoid bias problems reported by some studies in the use of partial eta square with small sample sizes (Larson-Hall, 2016). Cohen's d effect sizes equal or greater than 0.2, 0.5 and 0.8 were considered small, medium and large, respectively (Cohen, 1988). A Pearson correlational analysis (appendix 2) was computed to explore the hypothetical relationship between the dependent variables, namely between complexity and accuracy performance areas, as Skehan (2009, 2014, 2015) suggested. To determine the strength of the correlation, Cohen's (1988) guidelines were followed: r = 0.10 to 0.29, r = 0.30 to 0.49 and r = 0.50 to 1.0 were considered small, medium and large, respectively.

5. Results

5.1. Complexity

5.1.1. Descriptive statistics

The mean scores and standard deviations for linguistic complexity are given in **Table 1**. For syntactic complexity, the means for the two measures – clauses per AS-unit and coordinate clauses per AS-unit – suggest that the production of both planners and non-planners was more complex in the simple version of the task. However the mean length of clause was higher in the complex task than in the simple task. The two measures of lexical diversity, Guiraud's Index and VOCD, also diverged in terms of results. The means of the former measure were higher in the complex task under both planning and non-planning conditions. The means of VOCD pointed to a slight decrease in the complex task when participants were given time to plan their performance, (53.95 vs. 53.73), but for the non-planners the decrease was more evident (59.42 vs. 54.82).

5.1.2. Inferential statistics

The results of the six repeated measures' ANOVAs on complexity are reported in **Table 2**. Task complexity yielded significant differences on the three measures of syntactic complexity: the number of clauses per AS-unit

Table 1: Descriptive statistics of the measures of complexity.

		TASK COMPLEXITY															
		SIMPLE								COMPLEX							
		+ PLANNING				− PLANNING				+ PLANNING				− PLANNING			
	MEASURES	*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI	
				LB	UB			LB	UB			LB	UB			LB	UB
COMPLEXITY	Words per Clause	6.19 (0.67)	0.15	5.87	6.51	6.17 (0.66)	0.15	5.86	6.48	6.40 (0.72)	0.17	6.05	6.75	6.68 (0.81)	0.18	6.30	7.05
	Clauses per AS-unit	1.36 (0.18)	0.04	1.27	1.45	1.45 (0.23)	0.05	1.34	1.55	1.30 (0.18)	0.04	1.21	1.38	1.26 (0.16)	0.04	1.18	1.33
	Coordination	0.08 (0.07)	0.01	0.5	0.12	0.06 (0.06)	0.01	0.02	0.09	0.05 (0.04)	0.09	0.03	0.07	0.04 (0.05)	0.01	0.02	0.06
	Guiraud	7.14 (0.83)	0.19	6.74	7.54	7.26 (0.92)	0.21	6.83	7.69	7.78 (0.65)	0.15	7.46	8.09	7.90 (0.76)	0.17	7.54	8.26
	VOCD	53.95 (14.59)	3.35	46.92	60.99	59.42 (16.20)	3.62	51.84	67.01	53.73 (13.79)	3.16	47.08	60.38	54.82 (10.34)	2.31	49.98	59.66

Table 2: Results of ANOVAs on the measures of complexity.

Measures		*F*(*df*)	*p*	η_p^2	Observed Power
Task complexity	W/Cl	15.50(1,37)	.000***	.30	.97
	Cl/AS	18.79(1,37)	.000***	.34	.99
	Co/AS	5.51(1,37)	.024*	.13	.63
	Guiraud	38.35(1,37)	.000***	.51	1
	VOCD	2.31(1,37)	.137	.06	.32
Planning time	W/Cl	.371(1,37)	.546	.01	.09
	Cl/AS	.19(1,37)	.662	.01	.07
	Co/AS	1.11(1,37)	.298	.02	.18
	Guiraud	.26(1,37)	.614	.02	.08
	VOCD	.62(1,37)	.435	.02	.12
Complexity × planning	W/Cl	2.65(1,37)	.112	.07	.35
	Cl/AS	4.47(1,37)	.041*	.11	.54
	Co/AS	1.18(1,37)	.285	.03	.19
	Guiraud	.001(1,37)	.970	.00	.05
	VOCD	1.90(1,37)	.176	.05	.27

Note: *$p < .05$, **$p < .01$, ***$p < .001$.
W/Cl = words per clause; Cl/AS = clauses per AS-unit; Co/AS = coordinate clauses per AS-unit.

decreased in the complex task, $F(1, 37) = 18.79$, $p =< 0.001$, d = 0.69; the number of coordinate clauses per AS-unit reached also significance but the effect size was small, $F(1, 37) = 5.51$, $p = 0.024$, d = 0.38. However, the clause length increased in the complex task with a medium effect size, $F(1, 37) = 15.50$, $p =< 0.001$, d = 0.51. Among the two measures of lexical diversity, Guiraud's Index reached significance with a large effect size, $F(1, 37) = 38.35$, $p =< 0.001$, d = 0.81. An interaction effect of task complexity and planning time was found with respect to the number of clauses per AS-unit, $F(1, 37) = 4.47$, $p = 0.041$, $\eta_p^2 = 0.11$.

5.2. Accuracy

5.2.1. Descriptive statistics

Table 3 presents the descriptive statistics of all the accuracy measures. In the case of the percentage of error-free clauses, planners showed a higher mean in the complex task (32.06 vs. 33.47), while under the non-planning condition, the mean of the complex task decreased, (28.05 vs. 24.67). The number of errors per 100 words confirmed these results: the mean of the simple task was higher than in the complex version (19.08 vs. 17.18), but without planning time the mean of the complex task was higher (simple task $M = 21.94$ and complex task $M = 22.18$). The means of lexical errors per AS-unit and omissions per AS-unit did not show obvious differences between simple and complex tasks. There was a small decrease in the mean of lexical errors for the complex task under the planning condition (0.32 vs. 0.35), but a small increase in the number of omissions (0.29 vs. 0.28). Under the non-planning condition, the mean of the lexical errors was almost the same (simple task $M = 0.43$ and complex task $M = 0.44$), and the omissions also slightly increased (0.27 vs. 0.31). The scores obtained in the case of morphosyntactic errors per AS-unit were lower in the complex version of the task for both planners and non-planners. Finally, the percentage of self-repairs per errors was larger in the complex task ($M = 20.97$) than in the simple task ($M = 15.70$) under planning conditions, but went in the opposite direction under the non-planners (in the simple task $M = 12.02$ and in the complex task $M = 11.46$).

5.2.2. Inferential statistics

Table 4 presents the statistics of the repeated measures' ANOVAs on accuracy. There were significant effects of increased task complexity in one specific accuracy measure, morphosyntactic errors per AS-unit, $F(1, 37) = 31.92$, $p =< 0.001$, d = 0.46. A main effect of planning time was detected in four measures of accuracy. Under the planning time condition, the participants produced significantly more accurate speech with regard to (i) the total errors per 100 words, $F(1, 37) = 4.30$, $p = 0.045$, d = 0.64; (ii) the lexical errors per AS-unit, $F(1, 37) = 4.18$, $p = 0.048$, d = 0.57 and (iii) the morphosyntactic errors per AS-unit, $F(1, 37) = 7.62$, $p = 0.009$, d = 0.82. The effect size was considered medium (d = 0.52) for the error-free clauses, although it did not reach statistical significance. Finally, planning time significantly affected the repair behaviour of the learners, $F(1, 37) = 8.24$, $p = 0.007$, d = 0.77. This was the only measure with a significant interaction effect between the two independent variables, $F(1, 37) = 4.40$, $p = 0.043$, $\eta_p^2 = 0.11$, as with pre-task planning time the percentage of self-repairs increased when participants engaged in the

Table 3: Descriptive statistics of the measures of accuracy.

		TASK COMPLEXITY															
		SIMPLE								COMPLEX							
		+ PLANNING				− PLANNING				+ PLANNING				− PLANNING			
	MEASURES	*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI	
				LB	UB			LB	UB			LB	UB			LB	UB
ACCURACY	% Error-free clauses	32.06 (13.60)	3.12	25.51	38.62	28.05 (12.0)	2.68	22.43	33.67	33.47 (14.5)	3.32	26.51	40.44	24.67 (9.34)	2.09	20.29	29.03
	Errors/ 100 words	19.08 (5.92)	1.36	16.23	21.94	21.94 (6.08)	1.36	19.09	24.78	17.18 (6.31)	1.45	14.14	20.22	22.18 (6.42)	1.44	19.17	25.18
	Lexical errors per AS-unit	0.35 (0.12)	0.03	0.29	0.41	0.43 (0.20)	0.05	0.34	0.53	0.32 (0.17)	0.39	0.24	0.40	0.44 (0.18)	0.04	0.35	0.51
	Omissions/ AS-unit	0.28 (0.10)	0.02	0.23	0.33	0.27 (0.14)	0.03	0.23	0.33	0.29 (0.13)	0.03	0.23	0.36	0.31 (0.14)	0.03	0.25	0.38
	Morphosyntactic errors/ AS-unit	0.89 (0.31)	0.07	0.74	1.03	1.16 (0.40)	0.09	0.97	1.35	0.69 (0.28)	0.06	0.56	0.82	1.00 (0.39)	0.09	0.82	1.18
	% of self-repairs/ total errors	15.70 (8.55)	1.96	11.58	19.81	12.02 (7.15)	1.60	8.68	15.37	20.97 (10.81)	2.48	15.76	26.17	11.46 (6.51)	1.46	8.41	14.51

Table 4: Results of ANOVAs on the measures of accuracy.

		F(*df*)	*p*	η_p^2	Observed Power
Task Complexity	%EF	.46(1,37)	.503	.01	.10
	E/100W	2.00(1,37)	.165	.05	.28
	LE/AS	.17(1,37)	.680	.01	.07
	OM/AS	1.92(1,37)	.175	.05	.27
	MSE/AS	21.92(1,37)	.000***	.37	1
	SR/TE	2.87(1,37)	.099	.07	.38
Planning Time	%EF	2.98(1,37)	.093	.08	.39
	E/100W	4.30(1,37)	.045*	.10	.52
	LE/AS	4.18(1,37)	.048*	.10	.51
	OM/AS	.01(1,37)	.914	.000	.05
	MSE/AS	7.62(1,37)	.009**	.17	.77
	SR/TE	8.24(1,37)	.007**	.18	.78
Complexity × Planning	%EF	2.70(1,37)	.109	.07	.36
	E/100W	3.32(1,37)	.076	.08	.43
	LE/AS	.34(1,37)	.566	.01	.09
	OM/AS	.64(1,37)	.430	.02	.12
	MSE/AS	.32(1,37)	.574	.01	.09
	SR/TE	4.40(1,37)	.043*	.11	.53

Note: *$p < .05$, **$p < .01$, ***$p < .001$.
% EF = percentage of error-free clauses per total clauses; E/100W = errors per 100 words; LE/AS = lexical errors per AS-unit; OM/AS = omissions per AS-unit; MSE/AS = morphosyntactic errors per AS-unit; SR/TE = self-repairs per total errors.

complex task, but under the non-planning condition it was higher when they performed the simple task.

5.3. Fluency

5.3.1. Descriptive statistics

Table 5 summarises the scores on measures of fluency. A comparison across the columns reveals similar patterns between simple and complex task performances. The means of the speech rate (both pruned and unpruned) decreased slightly in the complex task under non-planning conditions. Under the planning condition, the mean of the pruned speech rate was higher in the complex task (45.52 vs. 44.96) but the mean of the unpruned speech rate was slightly lower (55.10 vs. 55.43). The means of fluency repair were larger in the simple version of the task for all participants.

5.3.2. Inferential statistics

Table 6 shows the results of the three repeated measures' ANOVAs on fluency. These results reveal that the two independent variables did not significantly affect the fluency of learners' output.

5.4. Perception of task difficulty

A subjective self-rating questionnaire (AVQ) was used to assess learners' perceived difficulty of task complexity, stress, confidence, interest and motivation. The results validated the manipulation of task complexity along the number of elements, $F(1, 37) = 7.42$, $p = 0.015$, d = 0.55, as the complex task was perceived as more difficult. The perception of difficulty was not significant along the ±planning time operationalisation. Concerning the amount of time given to plan the task ('I had time'/'did not have time to plan the task'), learners' perception was significant along the number of elements, $F(1, 37) = 6.29$, $p = 0.017$, d = 0.34, and along the planning condition, $F(1, 37) = 7.75$, $p = 0.008$, d = 0.79. Learners' perception of confidence and interest reached statistical significance: the level of confidence decreased along the complex task (more elements), $F(1, 37) = 7.74$, $p = 0.008$, d = 0.53, and learners showed more interest when performing tasks under non-planning conditions, $F(1, 37) = 7.04$, $p = 0.012$, d = 0.72. Perceived stress and motivation did not yield significant results.

6. Discussion

In the previous section, the statistical results are reported in detail. Here, the findings are interpreted and discussed in relation to this study's research questions and hypotheses.

> RQ1: Effects of increasing the cognitive task complexity along the factor ±planning time on the oral production of Chinese learners of PFL.

It was predicted that removing pre-task planning time would have negative effects on all dimensions of L2 oral performance. Concerning the manipulation of task complexity along planning time, the impact of this

Table 5: Descriptive statistics of the measures of fluency.

		TASK COMPLEXITY														
		SIMPLE								COMPLEX						
		+ PLANNING				− PLANNING				+ PLANNING				− PLANNING		
	MEASURES	*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI		*M* (*SD*)	*Std. Error*	95% CI
				LB	UB			LB	UB			LB	UB			LB	UB
FLUENCY	Speech Rate A	55.43 (8.56)	1.97	51.30	59.56	55.56 (13.64)	3.05	49.17	61.95	55.10 (9.50)	2.18	50.52	59.68	52.18 (11.61)	2.60	46.75	57.62
	Speech Rate B (pruned)	44.96 (9.19)	2.11	40.52	49.39	43.47 (10.89)	2.43	38.39	48.56	45.52 (8.55)	1.96	41.40	49.64	41.89 (9.51)	2.13	37.44	46.35
	Fluency repair	6.98 (3.00)	0.69	5.53	8.43	8.60 (3.32)	0.74	7.04	10.16	6.65 (2.79)	0.64	5.30	8.00	7.80 (3.37)	0.76	6.23	9.39

Table 6: Results of ANOVAs on the measures of fluency.

Effects on Different Measures of Fluency		*F*(*df*)	*p*	η_p^2	Observed Power
Task Complexity	Rate A	1.43(1,37)	.240	.04	.21
	Rate B	.16(1,37)	.694	.004	.07
	Repair	3.98(1,37)	.053	.10	.49
Planning Time	Rate A	.19(1,37)	.664	.01	.07
	Rate B	.84(1,37)	.366	.02	.15
	Repair	2.07(1,37)	.159	.05	.29
Complexity × Planning	Rate A	.96(1,37)	.333	.03	.16
	Rate B	.71(1,37)	.407	.02	.13
	Repair	.67(1,37)	.420	.02	.13

Note: $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$.

variable on accuracy reached statistical significance on four measures; in addition, although the percentage of error-free clauses was not significant, there was a medium effect size. Mochizuki and Ortega (2008) also reported gains in accuracy under five minutes of guided planning, but in the present experiment planning was unguided. Contrary to previous research (Ortega, 1999; Skehan & Foster, 1997; Wang, 2014), there were no significant effects of planning time on complexity and fluency. These results could be related to the amount of time allocated to plan the task. Following Levkina (2008), Levkina and Gilabert (2012) and Sasayama and Izumi (2012), learners were only given five minutes for strategic planning, although several studies (as mentioned previously) allocated ten minutes. If participants had been given more time to plan what to say and how to say it, the impact of this variable might have been more evident. Mehnert's (1998) work on the effects of different amounts of planning (one, five and ten minutes) showed that improvements in the different areas of performance (fluency, lexical density, accuracy and syntactic complexity) were only found in the production of learners given ten minutes for planning. It seems that in the current experiment planners prioritised form and thus they focused their attention on the formulation stage instead of using the time to organise their ideas, that is, concentrating on the conceptualisation of the message. The gains in accuracy did not result in trade-off effects with the other production dimensions, as they were not negatively affected. The stated hypothesis was only partially confirmed, as performing the task without strategic planning only decreased accuracy.

> RQ2: Effects of manipulating the cognitive task complexity along the variable ±few elements on the oral performance of Chinese learners of PFL.

It was hypothesised that increasing the number of elements of a task would decrease fluency but have a positive impact on accuracy and linguistic complexity. Contrary to other studies (Levkina, 2008; Levkina & Gilabert, 2012; Robinson, 2001; Sasayama & Izumi, 2012) that reported a decrease in fluency, in the current experiment there were no effects of task complexity (+few elements) on fluency. Concerning accuracy and linguistic complexity, the findings partially confirm the CH, as increasing the number of elements of the task led to less morphosyntactic errors per AS-unit, longer clauses and more lexical diversity. Michel (2011) also reported that the manipulation of the factor ±few elements resulted in more lexical diversity; however, this change was considered more quantitative than qualitative, as it could be explained by the input given (according to the author, learners used the words given in the input and not their own linguistic resources). In the present study the input was mainly visual, to avoid ambiguity in the interpretation of the results; nevertheless, of the two lexical diversity measures used, that is, Guiraud's Index and VOCD, only the former reached statistical significance. Malicka (2014) also found different findings for Guiraud's Index and D, a measure computed by the program D_Tools but essentially the same as VOCD (Meara & Miralpeix, 2017). In Malicka's (2014) study, only D reached statistical significance. Guiraud and VOCD were proposed to reduce the impact of text length, but their reliability is an ongoing discussion that is beyond the scope of this paper. Nevertheless, note that significant positive correlations were found between Guiraud and VOCD (i) in the simple task under planning time ($r = 0.84$; $p < 0.001$), (ii) in the simple task under non-planning time ($r = 0.75$; $p < 0.001$), and (iii) in the complex task under planning time ($r = 0.65$; $p < 0.01$). In the complex task without planning time, the effect size was medium ($r = 0.37$). These results suggest that, at the individual level, when Guiraud increased, VOCD was also higher.

Concerning syntactic complexity, the results for clause length and amount of subordination and coordination were different. In fact, performing the task with more elements resulted in gains in terms of words per clause, but both subordination and coordination lowered. The task type may explain these findings. As mentioned earlier, in previous research the manipulation of task complexity along the factor ±elements implied a number of options or aspects to take into consideration when taking a decision. The

argumentative task type was privileged by most researchers, except for Oh and Lee (2012) and Sasayama and Izumi (2012), who chose a narrative task. Performing a decision-making task implies that learners have to give reasons, justifications or opinions; increasing the conceptual demands of a task may trigger the use of specific and more complex language, such as subordination or connectors.

In the current study, a task involving giving-information was chosen, and perhaps the increased task complexity resulted in more lexically diverse output and longer clauses, as at the phrasal level learners used more words (for example, modifiers) to distinguish and refer to more elements. However, it also led to less subordination and coordination, as the task did not demand reasoning, in the same way as argumentative tasks. This study underscores the multidimensional aspect of linguistic complexity and its relation to task type, as length, variation and interdependence generated distinct results. Choosing different measures for this construct seems to be important if we want to have a more complete understanding of its subcomponents. If the decrease in coordination and subordination measures was due to the task's lack of reasoning demands, the improved output (in terms of clausal length, lexical diversity and lower morphosyntactic errors) suggests that performing the task with a greater number of elements directed learners' attentional resources towards both accuracy and complexity, or at least to some subdimensions of accuracy and complexity, as expecting improvements in all measures seems unrealistic. This explanation is supported at the individual level, as the correlations between accuracy and complexity measures under planned conditions in the complex task (+few elements) showed medium effect sizes, although without statistical significance. Words per clauses correlated with self-repairs ($r = 0.42$). The effect sizes were also medium for correlations between clauses per AS-unit and two accuracy measures (error-free clauses [$r = 0.38$]; and self-repairs [$r = 0.38$]). Clauses per AS-unit correlated negatively with errors per 100 words [$r = 0.34$]. These findings suggest that when learners produced more accurate speech, they also produced longer clauses and more subordination, and there were no trade-off effects. In the complex task without planning time, the words per clause negatively correlated with: errors per 100 words ($r = -0.59$; $p < 0.01$) and omissions per AS-unit ($r = -0.51$; $p < 0.05$). These results showed that when learners produced longer clauses their output was also more accurate. Regarding the relationship between the subordination measure and morphosyntactic errors the findings are different, as there were positive correlations between the clauses per AS-unit and the morphosyntactic errors per AS-unit ($r = 0.67$; $p < 0.01$). More studies with different task types will probably clarify these issues and bring new insights to this discussion.

RQ3: Interactions between the number of elements and the amount of pre-task planning time.

Based on the SSARC model, it was hypothesised that the attention-directing effect of performing a complex task along the number of elements would be boosted by decreased complexity along the resource-dispersing dimension (+planning). If so, the oral output of planners in the complex task (+elements) would exceed other groups in terms of qualitative changes (accuracy and complexity). The interaction results show effects between these two independent variables on one accuracy measure (percentage of self-repairs per total errors) and one complexity measure (clauses per AS-unit). Participants' self-repair behaviour increased when they engaged in the complex task with planning time, meaning that learners were directing their attention to form. Under non-planning time, increasing the number of elements of the task led to a lower percentage of self-repairs and thus less accuracy. In terms of complexity, learners produced fewer clauses per AS-unit when they engaged in the complex task under both planning and non-planning conditions, but non-planners' output was even less complex. The interaction effects found were only for these two measures, so these findings do not clearly confirm the potential synergistic effects of resource-directing and resource-dispersing variables, as predicted by Robinson's SSARC model for task design and sequencing. This issue remains under-researched and deserves future longitudinal investigation for a more clear-cut understanding.

7. Conclusion, limitations and future research

The aim of the current study was to investigate the impact on the oral production of Chinese learners of PFL of manipulating cognitive task complexity along pre-task planning time and the number of elements. The evidence showed that (i) planning time had an impact on accuracy and (ii) increasing the elements of a task positively affected one accuracy measure (morphosyntactic errors), lexical diversity (Guiraud's Index) and clause length, but coordination and subordination were negatively affected, which was explained by the task type chosen (an information giving task). These findings partially support Robinson's framework, and underlined the importance of using multiple measures for oral performance dimensions, to capture different patterns of the same construct. Finally, the importance of measuring the L2 oral performance in other languages (than English) should be emphasised. Considering the strong inflectional morphology of Portuguese and the syntactic differences between Portuguese and English, these results can bring new insights to the field of ISLA. Like all studies, this work has some limitations due to technical, human and time constrains, namely (i) a small sample size and (ii) the lack of lexical sophistication and breakdown fluency measurement. Lexical sophistication could not be measured because of the lack of available computerised tools for the Portuguese language. Regarding breakdown fluency, as it would have required measuring the number, duration and location of pauses' boundaries (Tavakoli, 2011), this dimension of fluency was not explored for practical reasons. In future research, such measures could give us a more complete picture of oral production in PFL. It would also be worth exploring the effects of the variables of this study (number of elements/planning time) on the oral performance of Chinese learners of PFL

with different task types, such as a decision-making task, to allow more precise comparison of results with previous studies. Additionally, prospective investigation on the effects of task complexity in connection with individual differences, learners' L1 different backgrounds (Cantonese vs. Mandarin), proficiency levels and task sequencing may inform decisions about syllabus and task design in PFL.

Additional Files

The Additional files for this article can be found as follows:

- **Appendix 1.** Example of tasks. DOI: https://doi.org/10.22599/jesla.40.s1
- **Appendix 2.** Correlations between accuracy and complexity measures. DOI: https://doi.org/10.22599/jesla.40.s2

Acknowledgements

I would like to thank the two anonymous reviewers for their suggestions and comments.

References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). USA: Erlbaum.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity and accuracy in L2 oral production. *Applied Linguistics*, *30*(4), 474–509. DOI: https://doi.org/10.1093/applin/amp042

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

Gilabert, R. (2005). *Task complexity and L2 oral narrative production*. Unpublished Doctoral dissertation. University of Barcelona.

Guará-Tavares, M. G. (2008). *Pre-task planning, working memory capacity, and L2 speech performance*. Unpublished Doctoral dissertation. Florianópolis: Federal University of Santa Catarina.

Guará-Tavares, M. G. (2011). Pre-task planning, working memory capacity, and L2 speech performance. *Organon*, *51*, 245–266.

Housen, A., Kuiken, F., & **Vedder, I.** (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.32

Kuiken, F., & **Vedder, I.** (2011). Task complexity and linguistic performance in L2 writing and speaking. In: Robinson, P. (ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, 91–104. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.2.09ch4

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: Housen, A., Kuiken, F., & Vedder, I. (eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, 143–170. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.32.07kui

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). New York: Routledge.

Levkina, M. (2008). *The effects of increasing cognitive task complexity along [+/− planning Time] and [+/− few Elements] on L2 oral production*. Unpublished Master dissertation. University of Barcelona.

Levkina, M., & **Gilabert, R.** (2012). The effects of task complexity on L2 oral production. In: Housen, A., Kuiken, F., & Vedder, I. (eds.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*, 171–198. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.32.08lev

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd Ed.) Mahwah, NJ: Lawrence Erlbaum Associates.

Malicka, A. (2014). *The role of task complexity and task sequencing in L2 monologic oral production*. Unpublished Doctoral dissertation. University of Barcelona.

Meara, P., & **Miralpeix, I.** (2017). *Tools for researching vocabulary*. UK: Multilingual Matters.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *SSLA*, *20*, 83–108. DOI: https://doi.org/10.1017/S0272263198001041

Michel, M. (2017). Complexity, accuracy, and fluency in L2 production. In: Loewen, S., & Sato, M. (eds.), *The Routledge Handbook of Instructed Second Language Acquisition*, 50–68. New York: Routledge.

Michel, M. C. (2011). Effects of task complexity and interaction in L2 performance. In: Robinson, P. (ed.), *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*, 141–174. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.2.12ch6

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. DOI: https://doi.org/10.1515/iral.2007.011

Michel, M. C., Kuiken, F., & **Vedder, I.** (2012). Task complexity and interaction (Combined) Effects on Task-Based Performance in Dutch as a Second Language. *EUROSLA*, *Yearbook*, *12*, 164–190. DOI: https://doi.org/10.1075/eurosla.12.09mic

Mochizuki, N., & **Ortega, L.** (2008). Balancing Communication and Grammar in Beginning-level Foreign Language Classrooms: A Study of Guided Planning and Relativization. *Language Teaching Research*, *12*(1), 11–37. DOI: https://doi.org/10.1177/1362168807084492

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

555–578. DOI: https://doi.org/10.1093/applin/amp044

Oh, M., & **Lee, H.** (2012). The effects of task complexity and task condition on learner language. *Korean Journal of Applied Linguistics*, *28*(4), 39–71. DOI: https://doi.org/10.17154/kjal.2012.12.28.4.39

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *SSLA*, *21*, 109–148. DOI: https://doi.org/10.1017/S0272263199001047

Révész, A. (2011). Task complexity, focus on L2 constructions, and individual differences: A classroom-based study. *The Modern Language Journal*, *95*, 162–181. DOI: https://doi.org/10.1111/j.1540-4781.2011.01241.x

Robinson, P. (2001). Task complexity, Task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics*, *22*(1), 27–57. DOI: 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. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *International Review of Applied Linguistics*, *45*, 1–32. DOI: https://doi.org/10.1515/iral.2005.43.1.1

Robinson, P. (2010). Situating and distributing cognition across task demands: The SSARC model of pedagogic task sequencing. In: Pütz, M., & Sicola, L. (eds.), *Cognitive Processing in Second Language Acquisition: Inside the Learner's Mind*, 239–264. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/celcr.13.17rob

Robinson, P. (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.2

Robinson, P. (2015). Cognition hypothesis, second language task demands and the SSARC model of pedagogic task sequencing. In: Bygate, M. (ed.), *Domains and Directions in the Development of TBLT*, 87–121. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.8.04rob

Sasayama, S. (2015). *Validating the assumed relationship between task design, cognitive complexity and second language task performance*. Unpublished Doctoral dissertation. University of Georgetown.

Sasayama, S. (2016). Is a 'complex' task really complex? Validating the assumption of cognitive task complexity. *The Modern Language Journal*, *100*(1), 231–254. DOI: https://doi.org/10.1111/modl.12313

Sasayama, S., & **Izumi, S.** (2012). Effects of task complexity and pre-task planning on Japanese EFL learners' oral production. In: Shehadeh, A., & Combe, C. A. (eds.), *Task-Based Language Teaching in Foreign Language Contexts: Research and Implementation*, 23–42. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.4.05sas

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

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency and lexis. *Applied Linguistics*, *30*(4), 510–532. DOI: https://doi.org/10.1093/applin/amp047

Skehan, P. (2014). *Processing perspectives on task performance*. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.5

Skehan, P. (2015). Limited attention capacity and cognition: Two hypothesis regarding second language performance on tasks. In: Bygate, M. (ed.), *Domains and Directions in the Development of TBLT*, 123–155. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.8.05ske

Skehan, P., & **Foster, P.** (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, *1*(3), 185–211. DOI: https://doi.org/10.1177/136216889700100302

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, *65*(1), 71–79. DOI: https://doi.org/10.1093/elt/ccq020

Tavakoli, P., & **Skehan, P.** (2005). Strategic planning, task structure, and performance testing. In: Ellis, R. (ed.), *Planning and Task Performance in a Second Language*, 239–273. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/lllt.11.15tav

Wang, Z. (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In: Skehan, P. (ed.), *Processing Perspectives on Task Performance*, 27–62. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.5.02wan

Wang, Z., & **Skehan, P.** (2014). Structure, lexis, and time perspective: Influences on task performance. In: Skehan, P. (ed.), *Processing Perspectives on Task Performance*, 155–185. Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tblt.5.06wan

Yuan, F., & **Ellis, R.** (2003). The Effects of pre-task planning and on-line on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1–27. DOI: https://doi.org/10.1093/applin/24.1.1

How to cite this article: Santos, S. (2018). Effects of task complexity on the oral production of Chinese learners of Portuguese as a foreign language. *Journal of the European Second Language Association*, 2(1), 49–62, DOI: https://doi.org/10.22599/jesla.40

Submitted: 12 January 2018 **Accepted:** 18 July 2018 **Published:** 31 August 2018

WHITE ROSE UNIVERSITY PRESS Universities of Leeds, Sheffield & York

Journal of the European Second Language Association is a peer-reviewed open access journal published by White Rose University Press.

OPEN ACCESS