International Journal of Chinese Language Teaching (2024)
Vol. 5 (2) 54-72 https://doi.org/10.46451/ijclt.20240206

Article

Chinese Second Language Learners' Speaking Development During Study Abroad: Complexity, Accuracy and Fluency

Rongrong Guo
University College Cork, Ireland

Received: 6 January, 2023/Accepted: 29 September, 2023/Published: 21 February, 2024

Abstract
The purpose of this study was to contribute to the literature on second language oral Complexity, Accuracy, and Fluency (CAF) development by assessing English-speaking learners of Chinese during Study Abroad (SA) in China, which have rarely been investigated in an Irish context. Moreover, relationships between the CAF constructs and those between the sub-constructs impacted by Study Abroad (SA) were discussed. Data were collected from ten English-speaking undergraduates of an Irish university from two curricular oral tests during pre- and post- 10 months' SA. Performance was elicited by topics, which were relative to the learning content when the learners were in the formal instruction context. To exhibit an in-depth evaluation of oral performance of instructed L2 Mandarin learners, fourteen CAF measures were analysed. The effects of SA on oral performance were explored by paired-samples t-tests. The results showed that the SA benefits oral gains in terms of speech fluidity, syntactic complexity (length and subordination), and lexical sophistication. Generalized from the analysis, trade-off effects are observed prevailingly between CAF constructs, while simultaneous improvements are present within CAF. This is attributable to the study abroad experience as well as the use of rehearsed monologue tasks in the study (Wright, 2020). It has shown that complexity and fluency were enhanced by pre-task planning (Skehan, 2009c; Skehan & Foster, 2001). Based on the findings, the study also provides pedagogical implications for the development of L2 Chinese oral performance in a university teaching setting.

Keywords
Chinese second language learners, study abroad, complexity, accuracy and fluency (CAF)

1. Introduction

Under the scope of oral assessment, using the Complexity, Accuracy, and Fluency (CAF) framework to assess L2 learners' oral performance is a growing area (e.g., Mora & Valls-Ferrer, 2012; Wright & Cong, 2014). The main factors, learning contexts, formal instruction (FI) at home and study abroad (SA), which affect oral development assessed by the CAF measures, as well as the relationship between CAF measures, have received sustained attention in the L2 field (e.g., Collentine, 2004;

Email: rguo@ucc.ie

Mora & Valls-Ferrer, 2012; O'brien et al, 2007). However, the effects of learning contexts on L2 speaking Chinese have only been assessed by a handful of studies (e.g., Du, 2013; Wright & Cong, 2014; Wright, 2018, 2020). Additionally, the effects of learning contexts on relationships between the CAF components and those between the sub-constructs within CAF in the L2 speaking Chinese have rarely been investigated. Therefore, the effects of learning contexts, in particular, study abroad on the oral CAF of adult English-speaking Chinese learners as well as on relationships between the CAF components and those between the sub-constructs within CAF will be analysed in the current study.

2. Literature Review

2.1 Speaking as a construct: Complexity, Accuracy, and Fluency (CAF)

The notions of complexity, accuracy, and fluency (CAF) can consistently capture components of L2 oral performance (Housen & Kuiken, 2009; Kuiken et al., 2019). Therefore, the CAF components are frequently used to assess L2 learners' oral performance (e.g., Housen et al., 2012; Skehan, 2003). For this reason, the CAF framework will be employed to assess L2 speaking Chinese in this study.

Complexity has been generally interpreted as the use of more challenging and difficult language and the extent to which learners can produce elaborate language (Ellis & Barkhuizen, 2005). There are two most components of linguistic complexity in L2 research have been widely analysed: syntactic complexity and lexical complexity. Accuracy refers to the extent to which an L2 learner's performance deviates from a norm (Housen et al., 2012), which is also termed correctness, coping with deviations from the norm, which are normally characterised as errors (Housen & Kuiken, 2009). Fluency is regarded as producing speech at the tempo of native speakers, which is not impeded by silent pauses, hesitations, filled pauses ('ers' and 'erms'), self-corrections, repetitions, and false starts (Lennon, 1990). Among the three constructs of CAF, fluency has received the most attention when measuring L2 Chinese learners' oral performance (e.g., Du, 2013; Feng, 2018; Wang, 2018; Wright & Cong, 2014; Wright, 2020).

Moreover, researchers have considered if and how these constructs of language performance interact. Concerning the correspondences between these three constructs, two theoretical hypotheses which aim to account for the impact of task type and task conditions on performance are most widely documented: Skehan's Limited Attentional Capacity model (Skehan, 1998a; Skehan & Foster, 2012) and the Cognition Hypothesis (Robinson, 2001, 2003, 2011).

2.2 L2 speaking development during SA

Study Abroad (SA) research has grown extremely rapidly over the last two decades, a situation which has been stimulated by the growing global popularity of SA programmes (Yang, 2016), and large-scale projects such as SALA (Perez-Vidal, 2014) and LANGSNAP (Devlin, 2019). It is traditionally assumed that L2 learners' language development is aided by extensive access to the target language during SA (Paige et al., 2012; Dewey et al., 2014). Specifically, overall proficiency has been revealed to significantly improve (e.g., Pérez-Vidal & Juan-Garau, 2011; Wright & Cong, 2014) alongside particular aspects of learners' linguistic development, such as oral fluency in terms of general fluidity (i.e., greater output, less silences) (e.g., Collentine & Freed, 2004; Du, 2013; LIanes & Serrano, 2017). Generalizing the findings of these previous SA research studies (e.g., Tullock & Ortega, 2017; Valls-Ferrer & Mora, 2014), it has been concluded that overall fluency increases during study abroad, as L2 speech becomes more rapid (speed), exhibits fewer and shorter pauses and hesitations (breakdown), and contains fewer self-repairs (repair).

Concerning the effects of SA on oral accuracy and complexity, there have been mixed findings. For instance, while some studies have shown accuracy to increase significantly during the SA context (e.g.,

Juan-Garau, 2014; Llanes & Muñoz, 2013; Pérez-Vidal et al., 2012), others have found no statistical improvement (e.g., Serrano, LIanes & Tragant, 2011). The discrepancy between the findings is very likely to result from the length of the SA period and the impact of the degree of immersion during the participants' stay in the target country.

Regarding lexical complexity, it seems there is a less clear benefit from SA. If there is any, this is generally attributed to the rich linguistic contact with native speakers that SA enables (Dewey, 2008). Applying CAF measures to assess lexical development, SA seems to be more advantageous for the development of oral production in terms of lexical diversity (measured by Guiraud's Index) (e.g., LIanes & Serrano, 2017; Juan-Garau & Pérez-Vidal, 2007; Mora & Valls-Ferrer, 2012). In contrast, some other research suggests that learners' lexical complexity does not significantly improve during SA (e.g., Pérez-Vidal & Juan-Garau, 2011; Serrano, LIanes & Tragant, 2011; Llanes & Muñoz, 2013; Wright, 2018, 2020).

In terms of syntactic complexity, several previous studies have found that SA is very beneficial (e.g., Juan-Garau & Pérez-Vidal, 2007; Jensen & Howard, 2014; Pérez-Vidal & Juan-Garau, 2011; Llanes & Muñoz, 2013; Mora & Valls-Ferrer, 2012) concerning overall complexity which is normally measured by the length of units. However, some other research has found no statistical significance concerning the complexity by subordination that is measured by the clauses per unit (LIanes & Serrano, 2017; Serrano, LIanes and Tragant, 2011; Mora & Valls-Ferrer, 2012).

Though contradictory findings have been revealed regarding the effect of SA on oral performance, a consensus has been reached that not all of the aspects of oral performance gain significant improvements. If gains are made, they tend to occur in oral fluency and vocabulary rather than accuracy and syntactic complexity (e.g., Leonard & Shea, 2017; Mora & Valls-Ferrer, 2012; Serrano, Llanes and Tragant, 2011; Valls-Ferrer & Mora, 2014; DeKeyser, 2014). This might result from the interaction between internal factors related to learners (e.g., the pre-departure proficiency; the age of participants) and external factors associated with the context (e.g., the duration of the SA programme).

2.3 L2 speaking Mandarin research in SA

The majority of SA research has focused on European languages. In particular, English and Spanish have become the two major target languages in this area (Yang, 2016). However, the effect of SA on L2 Mandarin has been under-explored and only a few empirical studies have been conducted in this area (e.g., Du, 2013; Kim et al. 2015; Wright & Cong, 2014; Wright, 2018, 2020). Moreover, there is an increasing trend in using a longitudinal design to measure L2 Chinese learners' oral development (Shi, 2002; Zhou, 2016; Wu, 2017; Wright, 2018, 2020) to complement the majority of SA research with a cross-sectional design. However, case studies are often applied due to the time-consuming nature of that type of research (Shi, 2002; Zhou, 2016). Therefore, it is necessary to examine the oral development of English-speaking learners of Chinese by employing CAF measures and a longitudinal design. This is particularly the case concerning the effects of study abroad on L2 Chinese speakers of English as this area has only been investigated by a handful of empirical studies (e.g., Du, 2013; Kim et al., 2015; Wright, 2020). Therefore, this study seeks to enrich this area by examining how studying abroad affects the speaking development of L2 learners of Chinese.

3. Study Rationale and Research Questions

This research aimed to explore the effects of study abroad on the oral development of English-speaking learners of Chinese by applying the CAF framework, and discussing the relationships between CAF constructs, and among the sub-components within CAF. The study compared the two widely applied competitive theories, the Trade-off Hypothesis (Skehan, 1998; Skehan & Foster, 1999) and the Cognition Hypothesis (Robinson, 2001, 2003). The study was orientated around two main research questions:

1. How do the complexity, accuracy, and fluency of the oral performance of instructed English-speaking L2 Chinese learners develop during pre- and post-SA?
2. Are the general relationships between CAF constructs, and the relationships between the sub-constructs of complexity and fluency, competitive or supportive in the oral performance of learners during pre- and post-SA?

4. Method

4.1 Participants

Followed appropriate ethical guidelines, the data analysed in this study were collected from 10 native English-speaking learners of Chinese who were undertaking an undergraduate degree in Commerce with Chinese Studies at a University in Ireland. No participants had contextual or long-term exposure to Mandarin before they studied the course. All of the students spent a full academic year abroad during year three, comprising approximately 10 months (from September to the following July), and were enrolled full-time at a university in Shanghai. During their time in China, the participants were expected to have comparable experiences based on reports of language usage declared by the programme manager of the host university.

The ten participants in this study were aged from 18 to 22 (See Table 1). There were more male participants (n=6) than females (n=4). In terms of the participants' pre-SA proficiency level, they were all enrolled in the same university degree and received approximately 360 hours of instruction in the Formal Instruction (FI) at home context. Furthermore, all the participants sat the HSK3 four months before the SA period to enable them to attain a one-year scholarship for the SA. However, three of them did not pass this exam (See Appendix). Their HSK levels were considered as their onset proficiency levels and were either HSK2 (n=3) or HSK3 (n=7). All learners were classified into upper beginner level considering the HSK 3 scores they achieved pre-SA (Appendix). When the participants reached their final attainment level one year after returning to the FI context, they had not reached HSK 4 considering the instructional level that they were allocated based on their performance. The participants were at lower intermediate level after the 10-month SA sojourn compared with when they were at upper beginner level pre-SA.

Table 1
The Profile of Participants

Participants No.	Age	Gender	Proficiency level (Pre-SA)
1	18	M	HSK3
2	18	F	HSK3
3	18	F	HSK3
4	22	M	HSK2
5	18	M	HSK3
6	18	F	HSK2
7	18	F	HSK3
8	19	M	HSK2
9	18	M	HSK3
10	18	M	HSK3

Note. The study will not take into account the difference in HSK levels between subjects in the data analysis portion. As a result, data at the group level will be provided ingoring individual performance at each level (HSK2 and HSK3).

4.2 Data collection

The participants were examined with a semi-longitudinal design that included two data collection times, Session 1 and Session 2 (See Table 2). The data were collected in a Formal Instruction at home (FI) context, which was before and after the 10-month SA period. During the FI at home period, there was not a particular focus on oral communication skills. Instead, the majority of the instruction consisted of traditional grammar teaching and practice, and learners experienced limited exposure to Chinese outside of the classroom. Data analysed in this study were speech samples elicited by topics, which were part of the continuous oral assessments within the curriculum of the college-level Chinese programme. A day before the students took the tests, they were able to access a larger number of fixed topics for preparation purposes. The topics were relative to the learning content when the learners were in the formal instruction context. However, none of the examination topics were revealed until they took the tests. During the oral tests, each participant was tested individually by their instructors. They were all asked to produce free speech.

Data collection were conducted in two sessions. In session 1,Students were tested at the end of the first semester of the second year of their course, prior to SA, thus enabling the study to investigate gains in acquisition starting from the baseline of the 360 hours of formal instruction that they had completed by this point in their degree programme. In session 2, students were tested three months after they returned to the FI at home context after their 10-month SA. During the SA period, the host college offered 540 hours of Chinese classes to the participants over two terms covering 34 weeks. The participants took the test when 3 months after return from China, at session 2, but only 5 weeks after restarting the year.

Table 2
Two Stages of Data Collection

Session	Session 1		Session 2
Date of oral tests	3-Dec		25-Oct
Year of students	Year-2	Year-3	Year-4
Semester	1		1
Hours of Instruction	360 hours	540 hours	36 hours
Context	FI	SA	FI

4.3 CAF analyses

To provide an in-depth evaluation of oral Complexity, Accuracy, and Fluency (CAF), fourteen measures were chosen to analyse the oral development of L2 learners of Chinese in this study.

Complexity was quantitatively analysed by lexical and syntactic complexity. Considering the variety in proficiency levels of the participants in this study, instead of applying type-token ratio, Guiraud's Index as an indicator of lexical diversity, was analysed. This approach intended to reduce the intervening effects of (oral and written) text length (Bulté & Housen, 2012). Lexical sophistication was measured by the ratio of words at different levels. Specifically, in this study, the new HSK (Hanban, 2012) was chosen as the corpus to categorise words at different levels. With regard to lexical sophistication, operationally, the words categorised under HSK 1 and 2 are considered as beginner level words, the words categorised

under HSK 3 and 4 are regarded as intermediate level, and the words categorised under HSK 5 and 6 and beyond are at the advanced level. The two most analysed syntactic complexity measures in previous studies are length and subordination (Kuiken et al., 2019). Therefore, these two indicators were also analysed in this study: 1) the number of syllables per AS-unit was calculated by the total number of syllables divided by the number of AS-units; 2) the number of sub-clauses per AS-unit was calculated by the total number of clauses divided by the total number of AS-units (e.g., Chen, 2015; Wu, 2017).

Lexical accuracy was obtained by the ratio of error-free lexical items, calculated by one minus the ratio of the lexical errors, which in turn was calculated by the number of lexical errors divided by the total number of lexical items. To measure the three sub-categories of utterance fluency (Skehan, 2003): speed fluency, and breakdown and repair fluency, the most commonly used indicators in each subcategory were employed. To assess speed fluency, Speech Rate (SR) and Mean Length of Runs (MLR) were used. SR was calculated by the total number of syllables (excluding filled pauses) divided by the time of utterances including pause time in seconds, multiplied by 60. This gave the produced syllables in one minute. MLR was calculated by the number of syllables divided by the number of silent pauses. To measure breakdown fluency, the silent pauses and filled pauses were coded and calculated separately. Two aspects were analysed in this study: the average length of pauses and the frequency of pauses. Specifically, four indicators were analysed: the average length of filled pause (ALFP), the average length of silent pause (ALSP), the number of filled pauses per 100 syllables (FP100), and the number of silent pauses per 100 syllables (SP100). Following Kormos (2006), repetitions, false starts, and self-corrections have been merged into the one category of dysfluency, which was assessed as a whole. The number of repetitions and repairs per 100 syllables (RR100) were analysed in relation to repair fluency.

5. Results

5.1 SA related results (S1-S2)

Prior to carrying out CAF analyses, data were tested for normality using the Kolmogorov-Smirnov test. All data were found to be normally distributed. A paired-samples t-test was conducted to compare the group means revealed by CAF in the pre-SA (S1) and post-SA (S2) conditions. The 14 CAF measures, categorised into the three domains: complexity, accuracy, and fluency related to the study abroad factor, are presented below.

Regarding the effect of the SA period on the oral performance of participants measured by the 14 CAF measures, the changes in performance from S1 to S2 can provide a picture of the difference between the group mean scores of CAF measures (see Table 3). Firstly, in terms of complexity and the two syntactic complexity measures, the length of AS-units showed a statistically significant increase (t = -6.034, $p < 0.001$) as did the subordination of AS-units (t = -9.432, $p < 0.001$). For lexical complexity, Guiraud's Index, which was used to measure lexical diversity, did not show any statistically significant difference (t = -9.432, p = 0.41). For lexical sophistication, beginner-level words decreased significantly (t = 4.387, $p < 0.001$), advanced-level words increased significantly (t = -7.665, $p < 0.001$), while there was a non-significant change in intermediate-level words (t = -1.291, p=0.114). The only measure of accuracy, lexical accuracy, showed a statistically significant decrease (t = 2.827, p = 0.01). Within the fluency domain, in terms of speed fluency measures, SR showed a statistically significant increase (t = -4.318, p <0.001). Similarly, MLR showed a statistically significant increase (t = -4.338, $p < 0.001$). For breakdown fluency, SP100 showed a statistically significant decrease (t =5.25, $p < 0.001$). Meanwhile, ALFP displayed a statistically significant increase (t = -2.157, p =0.03). The other two breakdown fluency indicators both revealed no significant differences - ALSP (t = 1.078, p = 0.155), and FP100 (t = 1.137, p = 0.142). The repair fluency measure, RR100, did not show a statistically significant change (t = 0.662, p = 0.262).

Table 3

SA Effects on Oral Performance (S1-S2)

Constructs			Indicators	Mean		t	p	Statistics
				S1	S2			
Complexity	Syntactic complexity		Syllables per AS –unit	18.11	34.04	-6.03	***.001	increase
	Clauses per AS–unit			1.50	2.09	-4.93	***.001	increase
	Lexical complexity	Lexical variety	Guiraud's Index	4.79	4.72	0.23	0.41	no difference
		Lexical sophistication	Lexical_ beginner	0.66	0.55	4.39	***.001	decrease
			Lexical_ intermediate	0.18	0.20	-1.29	0.114	no difference
			Lexical_ advanced	0.16	0.25	-7.67	***.001	increase
Accuracy	Lexical accuracy		Ratio of error-free lexical items	0.89	0.86	2.83	**.01	decrease
Fluency	Speed fluency		SR	91.33	113.20	-4.32	***.001	increase
	MLR			3.27	4.91	-4.34	***.001	increase
	Breakdown fluency		ALFP	0.49	0.56	-2.16	**.03	increase
			ALSP	0.89	0.81	1.08	0.155	no difference
			FP100	13.23	11.35	1.14	0.142	no difference
			SP100	32.00	21.69	5.53	***.001	decrease
	Repair fluency		RR100	3.13	2.79	0.66	0.262	no difference

Note. * $p < .05$; ** $p < .01$; *** $p < .001$

Speech rate (SR), mean length of runs (MLR), the average length of filled pause (ALFP), the average length of silent pause (ALSP), the number of filled pauses per 100 syllables (FP100), the number of silent pauses per 100 syllables (SP100), the number of repairs and repetitions (RR100)

5.2 Correlations between CAF constructs related to SA effects (S1-S2)

This section outlines the results of the correlations between subconstructs within complexity and fluency as well as the correlations between CAF constructs related to the effect of SA (S1-S2). The results can be explored by reviewing the data presented in Table 3.

As shown by the data, post-SA (S2), in general, within the complexity domain, the two syntactic complexity measures, complexity via length and subordination, had significantly improved compared to pre-SA (S1), suggesting a strong joint improvement after the SA. Meanwhile, for the indicators of lexical sophistication, at S2, advanced-level lexical items were significantly higher than S1 with a significant decrease in beginner level words, revealing significant growth in lexical sophistication. In

terms of Guiraud's Index, used to measure lexical diversity, no significant difference was observed between S1 and S2. Therefore, a weak trade-off effect was observed between lexical diversity and lexical sophistication.

A similar picture emerged for the fluency measures. Within the fluency domain, in general, at post-SA (S2), the results demonstrated a trade-off effect between speed and breakdown and repairs. The two speed fluency measures used in this study (SR and MLR) were both significantly higher in S2 than S1, suggesting a connected improvement. However, breakdown fluency measures did not yield a unified trend. Specifically, while ALFP at S2 was significantly higher than at S1, suggesting decreasing fluency, SP100, in contrast, at S2 was significantly lower than at S1, revealing increasing fluency. In this regard, the trade-off effect was evident within breakdown fluency. The other three breakdown fluency measures (ALSP, FP100, RR100) at S2 were not statistically different from S1. For repair fluency, no statistical difference was observed between S1 and S2. Broadly speaking, speed fluency measures gained great improvement, but repair fluency measures did not. Breakdown fluency measures showed moderate improvement.

Concerning the relationship between complexity, accuracy, and fluency, the analysis does not provide a unified picture, suggesting a trade-off effect at post-SA (S2). The majority of complexity measures, both syntactic complexity via length and subordination and lexical sophistication showed a joint improvement, whereas lexical variety, as measured by Guiraud's Index, did not change statistically. Within fluency, speed fluency measures (SR and MLR) revealed significant improvement. In contrast, within breakdown fluency, only ALFP and SP100 revealed a significant change. The other breakdown fluency measures (ALSP, FP100), as well as repair fluency, showed no difference. Lexical accuracy was observed to be significantly lower post-SA than pre-SA. This mixed picture will be discussed in terms of the trade-off effect that results from processing capacity limitations and task design.

6. Discussions

6.1 SA effects on oral performance (S1-S2)

Oral development measured by the 14 CAF measures was reported to explore the SA effects. The results are discussed below in terms of complexity, accuracy and fluency development during the pre- and post-SA periods.

6.1.1 Development of complexity

Syntactic complexity. Concerning syntactic complexity, the AS-unit length increased substantially after the 10-month SA period. The significant improvement after SA in this study is in line with previous studies (Jensen & Howard, 2014; Mora & Valls-Ferrer, 2012; Valls-Ferrer, 2010). This confirms that words per AS-unit as a measure of overall complexity in oral production clearly benefited from the SA period (Jensen & Howard, 2014; Juan-Garau & Pérez-Vidal, 2007; Mora & Valls-Ferrer, 2012).

The other indicator of syntactic complexity used in this research, the number of clauses per AS-unit to measure complexity via subordination, was also shown to benefit from the SA period. This is consistent with previous studies (Pérez-Vidal & Juan-Garau, 2011; Llanes & Muñoz, 2013) showing that the subordination of complexity achieved significant gains during SA. Additionally, the number of clauses per AS-unit has been proven to increase together with the proficiency level of L2 learners (Kuiken & Vedder, 2012). In this study, this also applied when learners were at lower intermediate level after the 10-month SA sojourn compared with when they were at upper beginner level pre-SA.

Lexical complexity. Concerning lexical complexity, lexical diversity as a subdomain measured by Guiraud's Index, did not reveal improvement during the SA period. This limited gain is in line with

previous studies (Pérez-Vidal & Juan-Garau, 2011; Wright, 2018, 2020) and further demonstrates that Guiraud's Index does not exhibit a significant increase within 10-month SA period. This finding further supports the notion that the development of lexical diversity is constrained by learners' proficiency levels. Specifically, advanced level learners outperform those at the beginner and intermediate levels and there is no significant improvement when learners are at the beginner and intermediate levels (Chen, 2015a; Ding & Xiao, 2016; Ye, 2015). Thus, in this study, lexical diversity did not show great gains pre-SA when the learners were at the upper beginner level and post-SA when they were at the lower intermediate level.

For lexical sophistication, after the 10-month SA, the beginner level (HSK 1 and 2) words decreased significantly and there was a significant increase in advanced-level (HSK 5 and 6 and beyond) words. Constrained by being at the lower intermediate proficiency level post-SA, the participants' intermediate-level (HSK 3 and 4) words did not show a statistical difference after the SA period. Meanwhile, the significant increase of advanced-level words after SA is largely attributable to the calculation method used in this research, which included the words in the HSK 5-6 level bracket and the words not included in the HSK. In particular, the study's data show that the participants acquired a large number of words beyond the HSK system during the SA. The significant increase of advanced-level words including non-HSK words, categorised as advanced level words, can be attributed to two key reasons. Firstly, during the SA sojourn, the participants accessed various types of input and more sophisticated words in their daily lives in the naturalistic environment, and their lexical repertoire consequently expanded. Indeed, it has been proven that vocabulary/lexical development can improve significantly because of an increased lexical repertoire during SA (Collentine, 2004; Milton & Meara, 1995; Jensen & Howard, 2014). Secondly, the textbooks that the learners used in the classroom setting in the college in China did not follow the HSK glossary. Therefore, during the formal instruction during SA, learners acquired a significant amount of non-HSK words. The HSK is widely used as a benchmark to assess lexical sophistication in existing Chinese studies, including this study. Therefore, advanced level words, including those words beyond the HSK system, increased significantly.

6.1.2 Development of accuracy

Accuracy was only measured with one sub-construct, namely lexical accuracy, and it saw a statistically significant decrease from pre- to post-SA. This is also in line with previous studies which showed that significant gains in oral accuracy after SA are not guaranteed (e.g., Mora & Valls-Ferrer, 2012 (3-month SA); Serrano, Llanes & Tragant, 2011(2-month SA); Valls-Ferrer & Mora, 2014 (3-month SA)), which very likely relates to the length of the SA period (2 to 3 months) in those studies. In this study, the learners did not make great gains in accuracy after a 10-month SA period, as accuracy was constrained by participants' proficiency levels during pre- and post-SA, that is, upper beginner to lower intermediate levels respectively. The results of previous research (i.e., Chen, 2015; Ye, 2015; Zhai & Feng, 2014) reveal that oral accuracy, in particular, lexical accuracy develops when learners are at the advanced level. No significant improvement can be expected when learners are at the beginner and intermediate levels. In this sense, constrained by the proficiency level of the participants in the research, no significant improvement could be expected when the participants were between upper beginner and lower intermediate levels.

6.1.3 Development of fluency

For the fluency measures, the results showed a speed fluency improvement (SR and MLR) as well as breakdown fluency (SP100). Both speech rate (SR) and mean length of runs (MLR) saw a significant increase after SA, meanwhile, the number of silent pauses per 100 syllables (SP100) decreased significantly after the 10-month SA. In other words, oral fluency showed speed improvement with fewer silent pauses. This is exactly in line with previous studies which assert that SA benefits oral fluency, in

particular, speed fluency (e.g., DeKeyser, 2014; Freed et al., 2004). After SA, learners are very likely to speak faster and they also produce longer speech runs and their speech becomes less hesitant, containing fewer pauses (Mora & Valls-Ferrer, 2012), in particular, silent pauses. The limited gains demonstrated in fluency breakdown and repair (i.e., ALSP, FP100, and RR100) were consistent with previous findings in that the participants did not show a significant decrease in dysfluency (i.e., filled pauses, mean length of pause, repairs, and repetitions) after the 10-month SA (Wright & Cong, 2014; Wright, 2020). The results are in line with previous research (Collentine & Freed, 2004; Mora & Valls-Ferrer, 2012; Valls-Ferrer & Mora, 2014) showing that learners are very likely to speak faster and that they also produce longer speech runs and their speech becomes less hesitant, containing fewer pauses. It can be concluded that after a 10-month SA period, fluency achieved significant gains and showed higher speed and longer speech runs with fewer silent pauses. Small and non-significant reductions were also found in the disfluency (total number of filled pauses) and repairs in the oral performance of English-speaking learners of Chinese.

Among the fluency indicators, SR, MLR, and SP100 were found to simultaneously improve after the 10-month SA. Referring to Levelt's speaking model, speech rate encompasses the working of the whole model, the conceptualiser, formulator and articulator (Towell et al., 1996), but particularly in formulation (Tavakoli & Wright, 2020). The significant increase of speech rate after SA suggests that the entire speech production process had been restructured, and that proceduralisation as a sign of increasing implicit acquisition of linguistic forms (Tavakoli & Wright, 2020) had occurred. As one temporal measure of fluency, mean length of run (MLR) has a conceptual connection with automatic speech production processing (Kahng, 2014), and has been suggested to be strongly associated with L2 fluency (e.g., Kormos & Denes, 2004; O'Brien et al., 2007). The increase in MLR is mainly attributable to the proceduralisation of different kinds of knowledge, including procedural knowledge of syntax and of lexical phrases. This might suggest that increased proceduralisation in the formulator of Levelt's speech model indicates greater time for planning each utterance and it should therefore become evident with longer pauses or a greater number of pauses (Skehan et al., 2016). This was supported by the significant decrease of the number of silent pauses (mostly occuring within AS-units observed in the speech samples in the study) per 100 syllables (SP100) after SA. However, filled pauses depends on where it is in terms of adding to or hindering fluency - if it's pre-clause it could be about message conceptualisation rather than utterance formulation repair; filled pauses can actually be a successful strategy.

As indicated above, apart from SR and MLR, another indicator of fluency that achieved significant improvement after the 10-month SA exposure to the target language context was SP100. The significant decrease of SP100 indicated improved fluency after the SA experience. This supports the hypothesis that silent pauses are a salient feature that determine speakers' fluency levels and contribute to judgments of nonfluency (Riggenhach, 1991). Interestingly, ALFP (the average length of the filled pause), as a breakdown fluency measure, saw a significant increase after SA. A significant increase of ALFP after SA might not be an indication of decreasing fluency. Rather, filled pauses can be used as a successful communication strategy for holding one's turn (Tavakoli & Wright 2020； Wright, 2020), and therefore may not be a clear indication of a lack of utterance fluidity (de Jong, 2016; Tavakoli, 2011).

Overall, with the exception of SP100 and ALFP, the dysfluency subconstructs, such as FP100 and repairs and repetitions, saw no statistical improvement from the SA experience. Levelt (1989) stated that speakers self-monitor their speech during the articulation stage with regards to any aspect of speech, such as content, syntax, choice of words, and phonological forms, and these are aspects which can be attended to simultaneously by native speakers. However, for L2 learners these processes are not yet automatised, which lead L2 speech to be more problematic (Kormos, 2006, 2011; Segalowitz, 2010). Gaps in linguistic knowledge or slow processing in accessing knowledge can impede the construction of accurate or sophisticated grammar and lexical items, resulting in reduced speech speed, hesitations, filled pauses, and repairs (Segalowitz, 2010, 2016; Tavakoli, 2011). Therefore, it is very likely that certain errors or dysfluency features can be attended to, while others might be ignored. Also, it has been suggested that both repairs and pauses act as monitoring processes during speech production, where the former is an

overt-monitoring process and the latter is a covert-monitoring process (Kormos, 2006; Tavakoli et al., 2016).

In conclusion, in this study, the benefits of the SA period mainly appeared in terms of significant improvements in the constructs of complexity and fluency, which are at the cost of accuracy. When learners produce speech at a higher speed, there are longer speech runs and fewer pauses. And their vocabulary becomes more sophisticated and their syntax becomes more complex, yet leading to more lexical errors.

6.2 Correlations between CAF Constructs related to SA effects (S1-S2)

This section firstly discusses the correlations between the subconstructs within complexity and fluency in the changes in oral performance of English-speaking learners of Chinese during the pre- and post-SA periods. This is followed by a discussion of the correlations between the CAF constructs.

6.2.1 Correlations between subconstructs within CAF after SA

Within the complexity domain, after the 10-month SA period, a strong joint improvement between lexical sophistication and syntactic complexity was clearly present, suggesting a supportive relationship within syntactic complexity. Within the construct of syntactic complexity, connected growth was evident between word complexity (average sentence length in morphemes) and sentence complexity (average number of clauses per sentence). These two indicators have been proved to be connected and supportive (Spoelman & Verspoor, 2010; Vercellotti, 2012, 2017, 2019). The growth processes of word complexity and sentence complexity are compatible with each other (Spoelman & Verspoor, 2010). An increase in clauses per AS-unit increases the overall length of an AS-unit (Vercellotti, 2012). There was therefore no evidence of a trade-off effect within the sub-constructs of syntactic complexity, which is consistent with previous research (Spoelman & Verspoor, 2010; Vercellotti, 2012).

Within the lexical subdomain of complexity, a weak trade-off effect was observed between lexical diversity and lexical sophistication. Lexical sophistication saw significant improvement, whereas lexical diversity had a non-significant difference compared to pre-SA. The development of lexical sophistication showed significant improvement after SA, especially advanced level words, including the non-HSK words that the learners acquired from their SA experience. However, the development of lexical diversity seemed to be related to the learners' proficiency level, supporting the findings of previous studies (Chen, 2015a; Ding & Xiao, 2016; Ye, 2015), that no significant improvement in lexical diversity can be expected when learners are at the beginner and intermediate levels. Moreover, the results can be understood in relation to Levelt's (1989) model of speaking. Lexical sophistication relates more to the conceptualiser stage of the model, whose output is the preverbal message. In contrast, lexical diversity is more closely related to the formulator stage, which accepts the preverbal message, and which then engages in processes of lemma selection and consequent syntax-building processes (Skehan, 2009a). For non-native speakers at lower proficiency levels, i.e., lower intermediate level after SA in this study, higher lexical sophistication (increased use of advanced level words) is more demanding in the conceptualiser stage following Levelt's model. This leads to negative implications in the formulator stage in terms of the retrieval of unusual lexical items. Thus, lexical diversity, as an indication of using unusual words, seems to have been impaired. Consequently, less demanding words were very likely to be produced more effectively. Retrieving more familiar words is easier because familiar topics create less challenges for L2 learners (Wright, 2020).

Similar to previous studies, correlations were examined between fluency measures (Mora & Valls-Ferrer, 2012; Tavakoli et al., 2016). The results, in general, showed a significant correlation in two aspects, which are described next.

Firstly, speed fluency, as measured by speech rate (SR) and mean length of runs (MLR), showed significant improvement. The significant increase of SR and MLR co-occured with a significant decrease in the number of silent pauses (SP100). This means that higher speed (higher fluency) and longer clusters of syllables between two pauses (higher fluency) co-existed with fewer silent pauses per 100 syllables (higher fluency), that is, the participants produced fewer silent pauses when producing longer utterances at a higher speed. Therefore, they improved their fluency in three ways. Furthermore, these results did not show a trade-off effect. There was a supportive relationship between speed fluency (i.e., SR, MLR) and breakdown fluency (i.e., SP100). This is consistent with the finding that the mean length of pause and mean length of the fluent run had weak to moderate negative correlations (Vercellotti, 2012) indicating a supportive relationship. This was very likely because of the benefit of the SA experience, which meant that learners were prone to speak faster and to produce longer runs with fewer silent pauses (Collentine & Freed, 2004; Mora & Valls-Ferrer, 2012; Valls-Ferrer & Mora, 2014).

Secondly, the analysis revealed a significant increase in ALFP, suggesting decreased fluency. This implies that longer utterances with higher speed containing fewer silent pauses (higher fluency) co-occurred with a longer length of filled pauses (lower fluency), which indicates a stretched syllable. These data support a trade-off effect between SR, MLR, SP100, and ALFP. Broadly, there was a tension between speed fluency and breakdown fluency in this regard. As indicated above, within the fluency domain, there was a competitive relationship between speed fluency (i.e., SR and MLR) and breakdown fluency (i.e., ALFP). For non-native speakers, filled pauses may be used as a successful strategy for holding one's turn during an utterance (Wright, 2020), because low-fluency speakers tend to use hesitations and non-lexical fillers to provide themselves with a longer period for processing (Levelt, 1989).

6.2.2 Correlations between complexity, accuracy, and fluency after SA

Generalising the analysis, the improvement in the majority of the complexity measures (syntactic complexity via length and subordination) and fluency, especially speed fluency, came at the expense of lexical accuracy with a longer length of filled pauses. Broadly, there was a trade-off effect between fluency and accuracy, following a "natural" meaning (fluency) -form (accuracy) tension predicted by Skehan (1998a) (Riggenhach, 1991). Moreover, a secondary tension within form, between control of form (accuracy) and interlanguage risk-taking (complexity) (Skehan, 1998b), was also observed.

The tension between complexity and accuracy observed in this study is unsurprising since an increase in complexity at the word and sentence level statistically increases the chances that more errors will occur. It is clear that an increase in complexity corresponds to a decrease in accuracy. In other words, more complex language is less likely to be error-free. This follows Skehan's (2009c) assumption of tension between control (accuracy) and risk-taking (complexity). These results imply that after a period of SA, when learners undertake rehearsed topic-prompted tasks, they are very likely to structure their language in a more ambitious manner. This "cutting-edge" language with more complex syntax and sophisticated words places significant demands on their attentional resources, and goes beyond what they can comfortably control. Therefore, accuracy becomes less controlled, leading to more errors (Foster & Skehan, 1996).

Between fluency and accuracy, a tension was revealed by a significant increase in speed fluency (SR and MLR) and a significant decrease in lexical accuracy. In other words, more lexical errors appeared with longer speech runs at a higher speed. Following the argument of Foster and Skehan (1996), that accuracy is concerned with form, learners attempt to maintain control over available resources and avoid mistakes in a more conservative manner instead of taking risks. Fluency reflects the primacy of meaning and the ability to communicate in real time. Fluency also prioritises idiom-based language over rule-based language to allow conversation to flow smoothly (Foster & Skehan, 1996). These results support Skehan's (1998, 2009c) theory that tension between focusing on meaning (fluency) and focusing on form

(accuracy) should be expected and that it will lead to a trade-off effect. As a result of the benefit from SA experience, as well as the effects of planning (Skehan, 2009c), learners seem to adequately produce idiom-based language to enable their utterances to proceed more smoothly (Foster & Skehan, 1996).

Connected improvement was broadly observed between complexity and fluency. Within the domain of complexity, length of AS-unit, calculated by the number of syllables per AS-unit, increased significantly (higher complexity) after SA. Similarly, the number of clauses per AS-unit achieved great gains (higher complexity). The same growth was also noted in lexical sophistication (higher complexity). Among fluency indicators, speed fluency (SR, MLR) saw a significant increase (higher fluency) with a significant decrease in SP100 (higher fluency). This joint increase in complexity and increased fluency are contrary to the trade-off effect. This is very likely attributable to two reasons. Specifically, SA experience advantages fluency (Freed et al., 2004; Mora & Valls-Ferrer, 2012) and complexity, especially syntactic complexity (Juan-Garau & Pérez-Vidal, 2007; Jensen & Howard, 2014; Llanes & Muñoz, 2013) and lexical sophistication (Collentine & Freed, 2004; Dewey, 2008; Kim et al., 2015). Moreover, because the learners undertook topic-promoted tasks with planning time in this study, complexity and fluency were promoted by planning time in general (Skehan, 2001, 2009c).

7. Conclusion

7.1 Summary of findings

The study sought to explore the oral CAF development of the same cohort of instructed English-speaking learners of Chinese and to investigate the interrelationships between the CAF constructs and the sub-constructs within CAF affected by 10-month SA. The main findings of the paired-samples t-tests for the first research question on oral development related to the study abroad are summarised below. During the pre- and post-SA periods, within the CAF constructs to measure oral performance, syntactic complexity (length and subordination) and lexical sophistication benefited significantly from the SA period. However, fluency only saw limited gains, but learners did produce longer fluent runs at a higher speed and fewer silent pauses. In contrast, accuracy decreased significantly, and the learners made more lexical errors after SA. This finding can be interpreted as showing that SA experience benefits complexity and fluency at the expense of accuracy. These findings are consistent with earlier research, which indicates that SA increases learners' oral fluency (e.g., Freed et al., 2004; Du, 2013; Mora & Valls-Ferrer, 2012; Trenchs-Parera, 2009; Valls-Ferrer, 2010; Wright & Cong, 2014) as well as syntactic complexity (Jensen & Howard, 2014; Juan-Garau & Pérez-Vidal, 2007; Mora & Valls-Ferrer 2012; Pérez-Vidal & Juan-Garau, 2011). Moreover, the task type has to be taken into consideration when interpreting learners' oral performance as measured by CAF. This study's analysis relates to the rehearsed topic-centered monologue task with planning that the participants undertook. It has shown that their complexity and fluency were enhanced by pre-task planning (cf. Skehan, 2001, 2009c).

Moreover, the generalised results of paired-samples t-tests for the second research question concerning the relationships between CAF constructs, and between subconstructs within CAF, and the impact of study abroad showed that:

The trade-off effect occurred between certain CAF constructs after SA, in particular between accuracy and complexity, and accuracy and fluency. These results confirm the Trade-off Hypothesis and that tension exists between control (accuracy) and risk-taking (complexity), and between focusing on meaning (fluency) and form (accuracy) (Skehan, 1998; Wang & Skehan, 2014). Task characteristics and learning contexts have been discussed to interpret the results because the different task and contextual characteristics supported different performance areas (Skehan & Foster, 2012). In terms of the learning context, study abroad favoured oral gains, especially fluency in terms of speed and silent pauses, syntactic complexity (length and subordination), and lexical sophistication. These findings broadly support Skehan's Trade-off Hypothesis, which postulates that increased performance in one area may drain the attentional resources available for other areas, leading to a potential decline in those areas'

performance (Skehan, 2009c). This study has also shown that accuracy suffers when complexity and fluency improve simultaneously. This is consistent with earlier empirical studies (Vercellotti, 2012, 2017), which looked at L2 English-speaking learners' performance on semi-spontaneous monologues with pre-planning time and found no evidence of a trade-off between complexity and fluency.

In conclusion, in this study, the trade-off effect was evident during the oral performance of English-speaking learners of Chinese. The trade-off effect was present not only between CAF constructs after SA (e.g., between complexity and accuracy) but also between subdomains within CAF after SA (e.g., between speed fluency and breakdown fluency). This contributes to research that on how learning contexts affect oral performance in L2 Chinese. However, some connected improvement occurred between CAF constructs, such as the joint improvement in complexity and fluency after SA. Likewise, connected improvement was also observed within certain subdomains of CAF. For instance, Within the complexity domain, a simultaneous improvement between lexical sophistication and syntactic complexity was observed after the 10-month SA period. Therefore, the trade-off effect was clearly important in the learners' oral development, while the exact pattern of the results that the learners achieved can be explained in relation to learning context (study abroad) and task design (rehearsed topic promoted monologues). These results contribute to research on the impact of study abroad on the oral performance of English-speaking Chinese learners by examining how attentional resources are prioritized across CAF dimensions and subconstructs within CAF.

7.2 Pedagogical implications

This research provides several pedagogical implications for English-speaking learners of Chinese and Chinese language teachers at college level in an Irish context. The challenging aspects (i.e., pauses, repairs and repetitions, lexical variety) of the oral performance of adult English-speaking learners of Chinese revealed by the CAF measures should be given more attention during teaching and practice within a teaching curricula. For example, to equip learners to have a better engagement during study abroad, oral class should be added as a transition between SA and FI contexts. To improve the oral performance of learners, oral fluency should be accorded more attention. Specifically, improvements in oral fluency can be achieved when dysfluency features (i.e., pauses, repairs and repetitions) are reduced.

Appendix: HSK Scores Per Participant

Participants No.	HSK3 (conducted on 28 March before study abroad)			
	Listening	Reading	Writing	Total
	100	100	100	300
1	90	61	75	226
2	88	77	63	228
3	68	51	67	186
4	73	44	55	172
5	83	47	55	185
6	63	47	59	169
7	90	74	83	247
8	68	47	47	162
9	80	77	71	228
10	88	61	51	200

Note. The version of HSK the participants took was "HSK 2.0", which was released in 2010. The passing score of HSK3 ("HSK 2.0") is 180.

References

Bulté, B. (2013). *The development of complexity in second language acquisition-A dynamic systems approach* (Doctoral dissertation, Free University of Brussels)

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21-46). John Benjamins Publishing Company.

Chen, M. (2015). Hanyu zuowei di er yuyan de langdu liuli du he zhunquedu de shiyan yanjiu. [An experiment study of reading fluency and accuracy in Chinese as second language acquisition]. *Hanyu yingyong yuyanxue yanjiu [Research on Chinese Applied Linguistics],* (1), 123-138.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*(2)*,* 227-248.

Collentine, J., & Freed, B. F. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition*, *26*(2), 153-171.

De Jong, N. H. (2016a). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 113-132.

DeKeyser, R. M. (2014). Research on language development during study abroad. *Language acquisition in study abroad and formal instruction contexts* (pp.313-325). John Benjamins Publishing Company.

Devlin, A. M. (2019). The interaction between duration of study abroad, diversity of loci of learning and sociopragmatic variation patterns: A comparative study. *Journal of Pragmatics, 146*, 121-136.

Dewey, D. P. (2008). Japanese vocabulary acquisition by learners in three contexts. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *15*, 127-148.

Dewey, D. P., Bown, J., Baker, W., Martinsen, R. A., Gold, C., & Eggett, D. (2014). Language use in six study abroad programs: An exploratory analysis of possible predictors. *Language Learnin*g, *64*(1), 36-71.

Ding, A. Q. & Xiao, X. (2016). Yidali Xuexi zhe chuji Hanyu kouyu cihui nengli fazhan yanjiu. [A study on the oral lexical development of elementary Italian learners' Chinese]. *Shijie hanyu jiaoxue [Chinese Teaching in The World], 30*(2), 239-252.

Du, H. (2013). The development of Chinese fluency during study abroad in China. *The Modern Language Journal*, *97*(1), 131-143.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language*. Oxford University.

Feng, Y. (2018). A case study on the diachronic development of oral Chinese fluency of primary Korean students, (Master's thesis, Jinan University).

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

Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition, 26*(2), 275-301.

Hanban (2012). New Chinese Proficiency Test (HSK) vocabulary (Revised Version)，Available on http://www.chinesetest.cn/godownload.do.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461-473.

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol. 32). John Benjamins Publishing.

Jensen, J., & Howard, M. (2014). The effects of time in the development of complexity and accuracy during study abroad: A study of French and Chinese learners of English. *Eurosla Yearbook, 14*(1), 31-64.

Juan-Garau, M. (2014). Oral accuracy growth after formal instruction and study abroad. *Language acquisition in study abroad and formal instruction contexts* (pp. 87-111). John Benjamins Publishing Company.

Juan-Garau, M., & Pérez-Vidal, C. (2007). The effect of context and contact on oral performance in students who go on a stay abroad. *Vigo International Journal of Applied Linguistics,* (4), 117-134.

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning*, *64*(4), 809-854.

Kim, J., Dewey, D. P., Baker-Smemoe, W., Ring, S., Westover, A., & Eggett, D. L. (2015). L2 development during study abroad in China. *System*, *55*, 123-133.

Kormos, J. (2006). *Speech production and second language acquisition*. Lawrence Erlbaum AssociatesKormos, J. (2011). Speech production and the Cognition Hypothesis. P. Robinson (Ed.) *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 39–60). John Benjamins Publishing Company.

Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, *32*(2), 145-164.

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. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 143-170). John Benjamins Publishing Company.

Kuiken, F., Vedder, I., Housen, A., & De Clercq, B. (2019). Variation in syntactic complexity: Introduction. *International Journal of Applied Linguistics*, *29*(2), 161-170.

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language learning*, *40*(3), 387-417.

Leonard, K. R., & Shea, C. E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal*, *101*(1), 179-193.

Levelt, W. J. (1989). *Speaking: From intention to articulation.* MIT press.

Liu, Y. (2017). Effect of task types on lexical complexity in L2 Chinese speaking performance. *Journal of Chinese Teaching in the World, 31*(2), 253-269.

Llanes, À., & Muñoz, C. (2013). Age effects in a study abroad context: Children and adults studying abroad and at home. *Language Learning, 63*(1), 63-90.

Llanes, À., & Serrano Serrano, R. (2011). Length of stay and study abroad: Language gains in two versus three months abroad. *Revista Española de Lingüística Aplicada*, 2011, núm. *24*, 95-110.

Llanes, À., & Serrano, R. (2017). The effectiveness of classroom instruction 'at home'versus study abroad for learners of English as a foreign language attending primary school, secondary school and university. *The Language Learning Journal*, 45(4), 434-446.

Milton, J., & Meara, P. (1995). How periods abroad affect vocabulary growth in a foreign language. *ITL-International Journal of Applied Linguistics, 107*(1), 17-34.

Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly, 46*(4), 610-641.

O'brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, 29(4), 557-581.

Paige, R. M., Berg, M. V., & Lou, K. H. (2012). *Student Learning Abroad: What our Students Are Learning, What they're Not, and What We Can Do About It.* Stylus Publishing, LLC.

Pérez-Vidal, C. (2014). Study abroad and formal instruction contrasted. *Language Acquisition in Study Abroad andFormal Instruction Contexts, 13, 17-58.*

Pérez-Vidal, C. & Juan-Garau, M. (2011). The effect of context and input conditions on oral and written development: A study abroad perspective. *International Review of Applied Linguistics in Language Learning, 49*(2),175–185.

Riggenbach, H. (1991). Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes, 14*(4), 423-441.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for examining task influences on SLA. In P. Robinson, *Cognition and Second Language Instruction* (pp. 287-318). Cambridge University Press.

Robinson, P. (2003). The cognitive hypothesis, task design, and adult task-based language learning. *The University of Hawai'I Second Langauge Studies Paper, 21* (2).

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

Robinson, P. (Ed.). (2011). Second language task complexity: Researching the cognition hypothesis of language learning and performance (Vol. 2). John Benjamins Publishing.

Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.

Segalowitz, N. (2016). Second language fluency and its underlying cognitive and social determinants. *International Review of Applied Linguistics in Language Teaching, 54*(2), 79-95.

Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Riggenbach (Ed.) *perspective on fluency* (pp.25-42). The University of Michigan Press.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*(2), 173-199.

Serrano, R., Llanes, À., & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System, 39*(2), 133-143.

Shi, J. (2002). Hanguo liuxuesheng Hanyu jushi xide de gean yanjiu [A case study of the acquisition of Chinese sentence patterns by Korean learners]. *Shijie hanyu jiaoxue [Chinese Teaching in The World], 4*, 34–42.

Skehan, P. (1998). Task-based instruction. *Annual review of applied linguistics, 18*, 268-286.

Skehan, P. (2001). Tasks and Language Performance Assessment. *In M. Bygate, P. Skehan, & M. Swain (Eds.), Researching Pedagogic Tasks Second Language Learning, Teaching and Testing* (pp. 167-185). Pearson Education.

Skehan, P. (2009a). Lexical performance by native and non-native speakers on language-learning tasks. In *Vocabulary studies in first and second language acquisition (pp. 107-124).* Palgrave Macmillan.

Skehan, P. (2009b). Models of speaking and the assessment of second language proficiency. In A. G. Benati (Ed.), *Issues in second language proficiency* (pp. 202-215). Continuum International Pub Group.

Skehan, P. (2009c). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30*(4), 510-532.

Skehan, P., & Foster, P., (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge University Press.

Skehan, P., and Foster, P., (1999). The Influence of Task Structure and Processing Conditions on Narrative Retellings. *Language Learning,* 49, 93-120.

Skehan, P., & Foster, P. (2012). Complexity, accuracy, fluency and lexis in task-based performance. *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 199–220). John Benjamins Publishing Company.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge University Press.

Skehan, P. Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics*, *54*(2), 97–111.

Skehan, P. (2003). Task-based instruction. *Language Teaching, 36*(1), 1-14.

Skehan, P. (1998a). Task-based Instruction. *Annual Review of Applied Linguistics, 18*, 268 – 286.

Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics, 31*(4), 532–553.

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching,* 54(2), 133-150

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, *65*(1), 71-79.

Tavakoli, P. (2016). Fluency in monologic and dialogic task performance: Challenges in defining and measuring L2 fluency. *International Review of Applied Linguistics in Language Teaching, 54*(2), 133-150.

Tavakoli, P., & Wright, C. (2020). *Second Language Speech Fluency: From Research to Practice*. Cambridge University Press.

Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics, 17*(1), 84-119.

Trenchs-Parera, M. (2009). Effects of formal instruction and a stay abroad on the acquisition of native-like oral fluency. *Canadian Modern Language Review, 65*(3), 365-393.

Tullock, B., & Ortega, L. (2017). Fluency and multilingualism in study abroad: Lessons from a scoping review. *System*, 71, 7-21.

Valls-Ferrer, M. (2010). Language acquisition during a stay abroad period following formal instruction: temporal effects on oral fluency development (Doctoral dissertation, Universitat Pompeu Fabra).

Valls-Ferrer, M., & Mora, J. C. (2014). L2 fluency development in formal instruction and study abroad. *Language acquisition in study abroad and formal instruction contexts* (pp. 111–136). John Benjamins Publishing Company.

Vercellotti, M. L. (2012). *Complexity, accuracy, and fluency as properties of language performance: The development of multiple subsystems over time and in relation to each other* (Doctoral dissertation, University of Pittsburgh).

Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics, 38*(1), 90-111.

Vercellotti, M. L. (2019). Finding variation: assessing the development of syntactic complexity in ESL Speech. *International Journal of Applied Linguistics, 29*(2), 233-247.

Wang, X. Z. (2018). A study on oral fluency of Chinese as a second language. (Doctoral dissertation, Northeast Normal University).

Wang, Z., & Skehan, P. (2014). Structure, lexis, and time perspective Influences on task performance. In P. Skehan (ed.), *Processing perspectives on task performance,* 155-185, Amsterdam, the Netherlands: Benjamins.

Wang, Z., & Skehan, P. (2014). Structure, lexis, and time perspective Influences on task performance. In P. Skehan (ed.), *Processing perspectives on task performance* (pp. 155-185). Benjamins.

Wright, C. (2018). Effects of time and task on L2 Mandarin Chinese language development during study abroad. In C. Sanz & A. Morales-Front (Eds.), *The Routledge handbook of study abroad research and practice* (pp. 166–180). Routledge.

Wright, C. (2020). Effects of task type on L2 Mandarin fluency development. *Journal of Second Language Studies, 3*(2), 157-179.

Wright, C., & Cong, Z. (2014). Examining the effects of Study Abroad on L2 Chinese development among UK university learners. *Newcastle and Northumbria Working Papers in Linguistics, 20*, 67-83.

Wu, X. Y. (2017). *A case study on the development of Chinese Proficiency of a Second-language Learner* (Master's thesis, Jinan University).

Yang, J. S. (2016). The effectiveness of study-abroad on second language learning: A meta-analysis. *Canadian Modern Language Review, 72*(1), 66-94.

Ye, W. (2015). Yingyu muyuzhe hanyu kouyu shuiping fazhan yanjiu [The oral Chinese language development of native English speakers]. *Nanjing Shifan Daxue Wenxueyuan Xuebao [Journal of School of Chinese Language and Culture Nanjing Normal University], 4*, 170–174.

Yuan, F. & R. Ellis. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 oral production. *Applied Linguistics, 24*(1), 1-27.

Zhai, Y., & Feng, H. (2014). Jiyu "kantu shuohua" renwu de Hanyu xuexizhe kouyu liulixing fazhan yanjiu [Study of Chinese learners' speaking fluency development with picture description activity]. *Huawen Jiaoxue yu Yanjiu [TCSOL Studies], 56*(4), 1–7.

Zhou, J. (2016). *A case study on the diachronic development of Korean students' oral Chinese accuracy and complexity* (Master's thesis, Jinan University).

Dr Rongrong Guo, Lecturer and coordinator, MA -Teaching Chinese to Speakers of Other Languages programme (MA-TCSOL), University College Cork, Ireland. Research interests: empirical studies in second language acquisition, language proficiency testing, teaching methodologies, and international Chinese education. Email: rguo@ucc.ie.

国际汉语教学学报（2024）
第 05 卷 · 第 02 期 54-72 https://doi.org/10.46451/ijclt.20240206

海外学习期间第二语言汉语口语的发展：复杂性、准确性和流利性

郭蓉蓉
科克大学，爱尔兰

摘要
本研究通过评估在中国留学期间英语母语者的汉语口语发展，为第二语言口语复杂性、准确性和流利性（CAF）发展的文献做出贡献。此外，本文还讨论了海外学习（SA）对 CAF 框架之间以及子框架之间关系的影响。数据来自爱尔兰一所大学 10 名学习汉语的英语为母语的本科生在 10 个月 SA 前后在与课程学习内容有关的两次口语测试的表现。数据结果表明，SA 在流利性、句法复杂性和词汇复杂性方面有利于口语习得。CAF 结构之间普遍存在权衡效应，而 CAF 内部存在同时进步的关系。这归因于被试的海外学习经历以及本研究采用的独白任务（Wright，2020）。此任务的任务前排练提高了被试口语的复杂性和流利度（Skehan，2001，2009c）。本研究也为大学阶段汉语作为二语口语的发展提供了教学启示。

关键词
汉语二语学习者，海外学习，复杂性、准确性和流利性

郭蓉蓉博士，爱尔兰科克大学国际汉语教育硕士项目（MA-TCSOL），讲师。研究兴趣包括二语习得的实证研究，语言能力测试，教学法，国际中文教育。