Book Section:

**Effects of Time and Task on L2 Mandarin Chinese Language
Development during Study Abroad
Authored by: Clare Wright**

In: The Routledge Handbook of Study Abroad Research and Practice

Doi:10.4324/9781315639970-11

Introduction

This chapter adds to the recent explosion of interest in second language (L2) Mandarin Chinese, by focusing on how linguistic proficiency develops during Study Abroad in China, across a range of written and oral tasks. The research presented here, drawn from one of the first systematic longitudinal investigations of Study Abroad (SA) in China (author and contributor), was to track written and oral proficiency development in L2 Mandarin Chinese, using an innovative task-based approach using both planned and unplanned written and spoken tasks. An exploratory approach was taken to explore why improvements in different tasks may be highly varied for participants, despite apparent similarity of their experience of immersion in the target language.

The focus on L2 Mandarin is particularly timely in SA research; given the rise in interest from students in learning Mandarin and institutions offering Mandarin classes either at home or via SA, there remain relatively few rigorous studies yet available (author1). Mandarin Chinese is also interesting for comparing written vs. oral language development, given its logographic non-transparent writing system. As SA has generally been found to aid oral proficiency more than other aspects of language, at least for languages with transparent writing systems, it is valuable to test this apparent trend in a Chinese setting to see how far the immersion experience affects development in written compared with oral proficiency.

We start with some key issues arising from SA research leading to an overview of the literature on SA in China, and then those few already published studies which are specifically related to language development in L2 Mandarin. Sources referred to use either Mandarin or Chinese to some extent interchangeably, but we use Mandarin here, to avoid confusion. We finish by evaluating empirical evidence from the study mentioned above, part of a wider case study project of ten students from UK visiting mainland China. The broader pedagogic and personal implications for students preparing to study abroad are also addressed, as well as indications of further research which is needed in this new burgeoning field.

Contextualising research into Study Abroad effects on L2 Mandarin

Research on L2 acquisition spans many disciplines – linguistic, cognitive, pragmatic, sociocultural, attitudinal, pedagogic, ethnographic, to name but a few – leaving our potential understanding of how different elements of SLA interact somewhat fragmented (author). Study Abroad (SA) research brings together many aspects of these sub-fields, particularly in relation to contextual effects on language development in adult L2 learners moving from instructed to

immersion settings. SA research thus offers the chance to resolve some of the fragmentation noted above, yielding crucial longitudinal or comparative data of how language knowledge and language use change, particularly in interaction, when the type of exposure to input changes. There is additional pedagogic value for SA research, in terms of evaluating the specific linguistic and cultural capital value of SA programmes, as the chance to travel abroad for study attracts more and more learners, and more language programmes are set up to promote SA (author).

SA is commonly assumed to offer a uniquely valuable environment to trigger significant improvements in language proficiency (Kinginger, 2011). However, the results of many current SA studies question this assumption (see, among others, Sanz, 2014; Mitchell, Tracy-Ventura and McManus, 2017), revealing much greater variability in SA outcomes than can be assumed. Some research demonstrate clear SA benefits for students' language proficiency compared to other contexts such as classroom teaching in the home country (e.g. Brecht, Davidson, & Ginsberg, 1995; Davidson, 2010; Du, 2013). Other studies (e.g., Collentine, 2004; Freed, Segalowitz & Dewey, 2004), by contrast, have found less clear effects of SA on language proficiency improvement; they claim that intensive home study can have more effect on certain aspects of language proficiency than SA, such as grammatical development. Collentine (2004) compared stay-at-home students with a comparable SA cohort, and found the at-home students improved more over a semester on a range of grammatical structures, such as tense and subordination. Isabelli-Garcia (2010) focused on acquisition of gender agreement, but similarly did not find differences in improvement between at-home students and SA students if measured over a semester. Another aspect of language proficiency is to examine SA effects on different language skills comparing writing and speaking, for example. Studies of written language development are not common in SA research, but recent work suggests that more than two semesters in the target setting may be required to show clear evidence of improvement (Sasaki, 2011; Serreno, Tragant & Llanes, 2012).

By comparison to the mixed findings on grammatical development and written proficiency, oral proficiency seems consistently to improve during SA (see, e.g., Collentine & Freed, 2004; O'Brien et al. 2007; Llanes & Munoz, 2009; Mora & Valls-Ferrer 2012). Yet even in SA oral studies, there remains a great deal of variability in outcome, sometimes explained as the SA experience not being as deep and effective as expected (Kinginger, 2011; Mitchell *et al.*, 2017), or due to differences in methods used. Precise operationalisation and assessment of L2 oral proficiency remains a debated area within SLA generally, let alone with SA research (de Jong, Steinel, Florijn, Schoonen, & Hulstijn,2012; Pallotti, 2009). Studies using generalised measures of proficiency such as the standardised Oral Proficiency Interview (OPI) are difficult to compare with studies using temporal measures, such as speech rate, pausing, length of speech run (e.g. as seen by comparing Freed *et al.*, 2004 with Brecht *et al.*, 1995). The role of quantity and quality of exposure in aiding development is also unclear: Du (2013) claimed that oral fluency is most influenced by time-on-task, i.e. the amount of time that students use the target language every day. However, Moyer (2013) found that quality of L2 experience is more important than quantity, as measured in terms of significant context-specific interaction. Meanwhile, author's (2013) longitudinal study of oral proficiency among 32 Mandarin learners

of English in the UK found no significant effect on improvement associated with qualitative or quantitative differences in target language use by the study participants.

Variability in SA findings is further compounded by methodological gaps or underinformativeness. Many studies are not designed for longitudinal comparisons of individual development; they may not control for differences in length of time spent abroad and level of exposure while abroad, or students' level of proficiency prior to SA. There is a growing awareness of the need to establish clear protocols for SA data collection (see Perez Vidal, 2014; Mitchell et al., 2017 for good examples of clarity). A range of tasks such as story-retelling and gap-fill may be used in SA studies (included in, for example the French Learner Language Oral Corpora, or FLLOC, Myles and Mitchell, http://www.flloc.soton.ac.uk/), making cross-study comparisons difficult. Additionally tasks themselves have variable inherent design effects (Skehan, 1998; Ellis, 2005; Robinson, 2001). Tasks can vary in how much planning and preparation time is allowed, or degree of cognitive load (Yuan & Ellis, 2003; Wang, 2014; author 3). All these factors can impact on the measure of proficiency which the task is aimed to capture; yet there seems to be currently little SA research which specifically mention task-related factors.

Finally, to date most of the published SA research has focused on the more widely spoken European languages, which can be argued to be linguistically, typologically and culturally reasonably close to each other (though of course to varying degrees). In order to advance our understanding of SA effects on language development, it is important to explore the issues outlined above relating to linguistic development, e.g. in grammatical knowledge, or writing or speaking abilities under different task conditions, when transferred to non-European languages, such as Mandarin.

Mandarin Chinese SLA in general, and in SA contexts specifically, addresses many interesting issues for linguistic research (see, e.g., Han, 2014), but also has pedagogic and practical implications, given current increased interest in studying Mandarin. Recent evidence from Chinese sources, such as the China Scholarship Council and the English-language Chinese University and College Admissions System (CUCAS), underscores the rapid recent rise in international students in China. By 2020, there are expected to be over half a million (China Scholarship Council, 2013), nearly doubling the recent levels of around 320,000 students over 180 countries, registered either for either degree or non-degree courses (CUCAS, 2013). Yet the explosion of interest in learning Mandarin remains relativley underexplored, particularly in specific terms of linguistic development, and somewhat disparately reported in SLA and pedagogic SA literature, at least in English-language publications.

There are new journals promoting Chinese research (such as Chinese as a Second Language Research Journal, and the Journal of Researching and Teaching Chinese as a Foreign Language), but they rarely address issues of linguistic interest to SA. Recent papers, Shi and Wen (2009, in Mandarin) and Zhao (2011), are useful overview of current findings relating to L2 Mandarin acquisition; however, these have focused on a limited range of specific linguistic features. More general publications on teaching and learning Mandarin (Han, 2014, Tao, 2016, and Lu, 2017) also have valuable contributions to our understanding of L2 Mandarin development, but do not particularly address the SA context.

General reports on SA progress that have included Mandarin learners may be based in individual institutions such as Georgetown University (VandeBerg, Connor-Linton, & Paige, 2009). VandeBerg *et al.*'s (2009) study referred to general progress assessed via a standardised ACTFL Oral Proficiency Interview (OPI), but did not report specifically on the improvement of the L2 Mandarin learners. Other SA studies on Mandarin learners have taken a broad interest in questions of sociocultural adaptation (Duff, Anderson, Ilnyckyj, Van Gaya, Wang, & Yates, 2013), or socio-affective and pragmatic development, such as Yu (2010) and Jin (2012). Yu (2010) collected questionnaires from 90 L2 Mandarin learners over a period of nine months including a period of SA. The study found some evidence that SA generated positive changes in attitudes and motivation and reduced language anxiety, matched by increased self-ratings on language proficiency. However, there was no objective measure of language proficiency reported in the study. Jin (2012) noted SA students' success in learning the use of compliment words, but did not report specifically on whether the students' language fluency or grammatical accuracy improved. Yang (2012) investigated patterns in L2 learners' pauses compared to native speakers, during a short pre-planned task, and found more pausing during clauses than at clause boundaries – similar to other comparisons of native/non-native speaker speech and pause patterns (e.g. Tavakoli, 2011), but lacked any longitudinal comparison to gauge SA effect.

In terms of grammatical accuracy or written linguistic development, very little seems to be published to date on Mandarin in Study Abroad settings, although work is emerging which has found SA benefits for specific constructions, such as the *ba*-construction (Du, 2016). Liu's (2009) study of Mandarin development on a SA programme takes a holistic approach integrating different tasks in different modes: OPI, Mandarin language standardised assessment tests, a portfolio of general writing tasks and a survey asking for self-ratings on reading, listening, speaking, writing, cultural awareness, and personal career development. The students all were able to reach advanced level on the OPI scores after fourteen months - two months' immersion at home, then eleven months' academic study and finally a month's residence in China. But the study does not detail which language features improved the most, at what point on the programme, or provide a breakdown of how oral and written linguistic development could be compared across tasks.

Du's (2013) study is one of only a handful that focuses on longitudinal change in L2 Mandarin speech in different tasks, finding significant improvements in both planned and unplanned output over one semester. Rich data was collected once a month for four months from a cohort of 29 students, using both recorded Chinese speaking classes for planned instructed output and using Labovian-style individual interviews to elicit spontaneous output. Other speech studies can be found in Mandarin (Chen & Wang, 2008; Chen, 2012; Sun, 2008; Zhao & Liu, 2013).

This study therefore adopts an exploratory approach, driven by the gaps in SA literature identified above, both in general terms and for L2 Mandarin specifically, assessing UK students' linguistic proficiency before and after two semesters' study in China. In order to give ecological validity to our research, students' actual examination test data from their university end of year summative assessments were used in a pre-posttest design, to get a

broad holistic picture of their language development. The findings discussed here thus aim to be an innovative exploration of longitudinal individual development of L2 Mandarin in both written and spoken data, in different tasks.

The Study

The study asked the following research questions:
(1)   How does written proficiency change after Study Abroad (SA)?
(2)   How does oral proficiency change after SA?

In answering these questions, we also addressed the issue of task effects on proficiency in planned or unplanned conditions (Ellis, 2005; Wang, 2014), predicting that most improvement would be seen in planned conditions.

Ten volunteers were recruited from a group of adult English university learners of Mandarin at a UK university; full institutional ethical procedures were followed. The participants (aged around 20 years old) had had two years' instruction *ab initio* at the UK institution, receiving typically six hours a week explicit language instruction. They had had no previous exposure to Mandarin nor had visited China. For their Study Abroad, the students went to one of five Chinese universities (Beijing, Shanghai, Xi'an, Chengdu, or Hainan) taking university language courses for two semesters, with formal class language instruction for twelve hours per week; three students stayed on after the end of the university classes for around a month for travel around China. All were in mixed-language residences while studying in China. Diary report data was collected on amounts of interaction out of class at the start of Semester 1, Time 1. Mean hours' daily interaction at Time 1 was reported as 1.45 hours (SD 0.84). A sub-group provided interaction reports mid-way through at time 2 (around five months) and at the end of Semester 2 at Time 3 (eight months). Mean interaction was reported at Time 2 as 1.87 hours (SD 0.54), and by Time 3 as 2.49 hours (SD 1.07). The participants were judged, as far as possible, to have had comparable experiences in exposure to Mandarin during SA.

Data were collected as a set of tasks in the context of the students' standard end of second year language assessments in their UK university, which were repeated after the period of study abroad, using the same assessed task procedures and setting, on the participants' return to the UK university at the start of year 4 (see author for further details). All students took all tasks at both times; twenty-two students were initially recruited, but only ten completed the full set of tasks at both times. We used three written and two monologic speaking tasks, part of the standard end of year test battery of coursework and examinations, divided into timed and untimed tasks, and further sub-divided into planned and unplanned tasks. The timed unplanned written tasks were i) a descriptive letter; and ii) a dialogue, both on the theme of daily routines or social activities, and completed as part of the students' end of year examinations. The untimed planned written task was a short paragraph on expectations of life in China, to be completed outside class, marked as coursework. The speaking tasks, timed to last around two minutes each, were taken from the oral section of the end of year examination. They were i) a planned talk on participants' daily life in China (thinking ahead to what they expected before

their study abroad at Time 1, or reflecting back after their return at Time 2); and ii) a description of a photo depicting a group of friends in a typical social setting. All the tasks therefore drew on common themes of daily life and social settings, designed to overlap on required grammatical structures and familiar lexis; however, the tasks were taken to reflect different degrees of planning allowed, and in the writing, any effect of time limit vs. no time limit, to allow us to tease out any evidence of specific task and mode effects.

Variables for the written tasks were as follows. The two timed tasks were assessed for accuracy, including grammatical, lexical and orthographic accuracy, and combined into a total % accuracy mark across both tasks. On the untimed task four specific linguistic variables were used: length (total characters), and specified morphemes to reflect increasing knowledge of grammatical complexity (*de*-possessive and *de*-relative), and use of *shi* copula to illustrate knowledge of discourse-level optionality (for a detailed rationale of the variables used, see author and author). The timed task data had been intended to be included in the analysis of length and target morphemes, but too little output was generated to allow this analysis.

For the oral tasks, we report here on five: output (total characters), lexical diversity (Guiraud's index or G, a type/token based measure), articulation rate (characters per second during speaking (scored from 0 to 1); phonation time ratio (what proportion of the whole time was spent speaking or in silence, scored from 0 to 1), and hesitation rate (numbers of filled pauses and repairs divided by total output, scored from 0 to 1). Pauses were set at 250ms, in line with standard practice in L2 fluency research (see e.g. author and other). These measures together represented key elements of oral proficiency – ease in accessing the knowledge base (total produced, lexical diversity), speech speed (articulation rate), and rate of breakdown (phonation time ratio, and hesitation rate). The oral files were transcribed using both CHAT (MacWhinney, 2000) and PRAAT softwares (Boersma & Weenink, 2014). For full details, see author.

There were a number of methodological issues and complex decisions to make in creating analysable transcripts. Firstly, there is a debate on what constitutes a Mandarin "word" (Du, 2013; Li & Yang, 2009, Yang & Sun, 2015); also many of the fluency studies on which this research was based use syllables rather than words in their calculations. In view of this, it was decided to transcribe the oral data as single characters. There are autosegmenting technologies available (e.g. in the Lancaster Los Angeles Corpus of spoken Chinese, Xiao & Tao, 2006); however, this provided problematic for our small corpus, which had a lot of "learner noise", requiring manual segmentation. There are potential challenges in reliability for our lexical measure taking a single character segmentation approach, but running an analysis on a sample of two participants' data, both with single characters and with manually identified two-character word boundaries where required, significant differences were not detected. Most of the characters in this corpus equated to a single monosyllabic item – the most frequent lexical items typically were *wo3* 我（"I"), *de* 的 (possessive marker), *shi4* 是 ("be"); *hen3* 很 ("very"), *hao3* 好("good"), *you3* 有 ("have"）. Therefore following other published work on L2 Mandarin (e.g. Du, 2013, 2016), it was decided that one-character segmentation would be a valid approach for this study. Secondly, this research is the first, to our knowledge,

to combine both PRAAT and CLAN methodologies in transcribing Mandarin – although this was not specifically because of analysing Mandarin, this approach was chosen as each package had analysis programs that made them ideal for calculating the specific measures examined here. Combining packages however created additional operationalisation challenges of matching transcription practice, to ensure temporal segmentation using PRAAT matched the start of utterance transcriptions in line with CHAT conventions. Here it was decided to transcribe and time the start of each utterance as initiated with some kind of lexically meaningful word, and where any following pause was less than 3 seconds. Finally, we were faced with the challenge specific to Mandarin of how to acknowledge whether repeated morphemes were part of hesitation (in terms of retracing and repair) or intensification, e.g. for the morpheme *zui4* 最("most") which when repeated means "very". To assure reliability and validity of the transcripts, we trained a team of two transcribers; a third researcher then inspected each transcript, producing an interrater reliability of 95%; all three transcribers then discussed any remaining differences till full agreement was reached.

Results

In response to Research Question 1, looking at development in grammatical accuracy in writing, including on three specified morphemes (*de*-possessive, *de*-relative and optionality of *shi* copula) we found little difference in on the timed tasks, but some change on the untimed task. Mean accuracy on the timed tasks (the descriptive letter and dialogue) remained similar. Mean combined scores were 70.61% (SD 10.27) at Time 1, and 71.9% (SD 11.6) at Time 2. There was also considerable individual variation – the highest score was above 80% before and after SA, while the lowest score was around 50% at both times (though this was not the same person). No changes were significant.

By comparison, the planned untimed piece of writing, the out of class essay, more clearly improved. In terms of total output, mean number of characters produced was 660.5 characters (SD 97.14) at Time 1, rising to 697 characters (SD 107.89) at Time 2, though this change was not statistically significant. Examining the specific target morphemes, there was a drop-off in use of *de*-possessive, from mean 20.38 (SD 6.34) to 9.45 (SD 4.37); there was a rise in use of *de*-relative, from mean 2.5 (SD1.38) to 18 (SD 8.17). The reduction in the *de*-possessive morpheme was statistically significant ($p < .001$, with a small Cohen's D effect size of 1.95). The increase in the *de*-relative was also statistically significant ($p < .001$, with a medium Cohen's D effect size of 5.7). The participants' grasp of the appropriate optionality of the *shi*-morpheme was relatively low at either time (Mean at Time 1 was 4.55, SD 2.25; Mean at Time 2 was 3, SD 1.86); the slight decrease was non-significant.

Turning to Research Question 2, regarding oral proficiency, comparing a planned or unplanned descriptive monologue task, the five key variables selected to index fluency development were: total characters produced, lexical diversity (G), phonation time ratio, hesitation rate, and articulation rate (see author and author for details on operationalisation for each variable).

All the variables on the planned task improved though not all significantly. The findings

are summarised in the table below, with effect size shown as Cohen's D for any significant differences.

Table 1: Oral proficiency planned monologue results

	Time 1 Mean (SD)	Time 2 Mean (SD)	Significance (Cohen's D)
Total characters	169.9 (44.2)	290.6 (101.28)	$p < .01$ (1.66)
Lexical diversity	5.56 (.96)	6.07 (.41)	ns
Phonation-time ratio	0.69 (.09)	.075 (.08)	$p < .05$ (1.66)
Hesitation rate	0.14 (.06)	0.13 (.05)	ns
Articulation rate	2.85 (.48)	3.09 (.26)	ns

As in the written data, there was evidence of individual variation with wide ranges on all measures, but in general individual variation reduced (in terms of smaller SD by time 2) on all measures apart from output.

For the unplanned task, again all the variables improved, and more of the changes were significant, though all the effect sizes were very small, as noted in the table below.

Table 2: Oral proficiency unplanned monologue results

	Time 1 Mean (SD)	Time 2 Mean (SD)	Significance (Cohen's D)
Total characters	120.7 (39.71)	174.4 (78.29)	$p <.01$ (1.08)
Lexical diversity	5.04 (.49)	5.67 (.66)	ns
Phonation-time ratio	0.59 (.11)	0.71 (.11)	$p <.05$ (.84)
Hesitation rate	0.24 (.11)	0.15 (.06)	$p <.05$ (.79)
Articulation rate	2.1 (.27)	2.7 (.89)	$p <.05$ (.84)

Again there was a wide individual range on all measures; on this task, SD only reduced on hesitation rate.

We noted that overall scores tended to be better across the spread of variables on the planned task at both times. However, the scores on the unplanned task at Time 2 seemed to improve more – i.e. the participants seemed to catch up in performance in the more difficult task by Time 2. We ran Analyses of Variance (ANOVA) on the Time 1 and the Time 2 scores, using Task as group factor to try to confirm this interpretation. We found a significant effect for Task ($F_{(1, 9)}$ = 9.952, p = .008, partial η^2 = .869) at Time 1, but no significant effect at Time 2. Paired t-tests between the task scores at Time 2 showed that participants were indeed not significantly different in performance on the majority of measures for both tasks at Time 2.

Discussion

Our research questions compared L2 development in Mandarin in written and in oral

tasks, and whether task factors – timed/untimed (for writing) or planned/unplanned (for speaking) had an effect. We found that SA overall had a clearer effect on the oral tasks than on the written tasks; also that that there were differences arising from the different tasks, most significantly in the oral data but also in the written data.

The written tasks and oral tasks were not experimentally comparable, due to the exploratory nature of the study design, which used students' standard end of year university assessments through coursework and timed examinations. Here written Mandarin development was evaluated, first, through holistic accuracy on two timed exam tasks (a letter and a dialogue); second, in an untimed coursework essay, through total amount written and also accuracy on specific morphemes – the *de*-possessive, *de*-relative morphemes, and the optionality of the *shi* copula. Due to insufficient data in the timed tasks, the target morphemes could only be analysed in the untimed task. Taken together, the evidence of accuracy (holistic or specific) and overall amount written were evaluated for improvement in written proficiency. However, we did not find much evidence of consistent improvement in the written tasks. Overall it seemed that writing in characters remained particularly challenging even after the period of SA, though there was considerable individual variation. Some participants from the total cohort failed to complete both timed writing tasks, and on the untimed task, some of the essays were markedly shorter than the others, even at Time 2. It was noted in the untimed coursework task that there was a greater grasp of the more complex *de*-relative morpheme, though this did not apparently carry through to more pressured timed tasks - in a post-hoc analysis of the timed tasks, this structure was not substantively more evident at Time 2 despite its higher occurrence in the untimed task.

Given the mixed evidence of improvements in written tasks, we suggest that further more systematic empirical evidence is needed to explore further whether the lack of consistent and evident improvement in writing was down to L2-specific problems (using characters, particularly under pressure of time), or amount of writing practice during the SA experience, or more generic problems in foreign language writing, particularly for certain tasks, as has been noted in the L2 English writing literature (e.g. Ferris & Hedgcock, 2005; Hamp-Lyons, 1991). Nevertheless we conclude that moving away from a generic assessment of progress measured simply in accuracy or total output in typical summative exam marking schemes obscures more detailed evidence of linguistic progress e.g. in using increasingly complex morphemes, and we recommend SA research to better integrate linguistic and pedagogic/assessment approaches, particularly given L2-specific issues in developing literacy in Chinese.

In terms of the second research question on oral development, as noted above, the study design was intended to capture detailed linguistic insight into how oral production changed over time, by using a wider range of variables than is often used in SA methodology and which might be obscured within standard summative exam marking scores. We also wanted to look for any task effect of planning time comparing the two spoken tasks, given the evidence form the written tasks of increased output on the untimed task (as we assume that untimed writing includes the benefit of planning time found on speaking tasks).

We found a mix of significant and non-significant (but evident) improvements on both the planned and unplanned tasks. This supports other SA literature that time spent immersed in

the target language environment has a clear beneficial effect, since increased opportunities to talk is argued to aid both utterance fluency in articulatory terms, but also in "cognitive fluency" (Segalowitz, 2010) across both planned and unplanned speech.

We also found a task-related difference, in that task differences at Time 1 were significant on all variables. This is not surprising, given the common benefits on fluency found if planning or advanced preparation time is given (Ellis, 2005). However, we noted that these differences had reduced or were no longer significant by Time 2 – in other words, immersion had triggered more development on the unplanned task than on the planned task (although greater variability in improvement was found at individual level in the unplanned task). We assume that Task 1, which the participants planned for beforehand, potentially allowed for participants memorise and recite their talk. This task factor created a performance advantage on many speech measures at Time 1, and also yielded more output at Time 2. We believe this is due to planning and rehearsal aiding automaticity in articulation, by reducing any need to construct meaning in real time. This is similar to de Jong & Perfetti's (2011) study, which found that repeating the same story again and again, but under increasing time pressure, resulted in greater temporal fluency in terms of smoother faster speech. This kind of rehearsed articulatory fluency equates to what we term "performative competence" (Author).

Task 2, requiring unplanned, spontaneous speech, created predictable higher hesitation rate and slower articulation rate but these were less significantly a problem by Time 2. We assume this indicates that the quality and quantity of exposure during SA created plenty of opportunities for participants to experience input and practice their spontaneous spoken output in meaningful interactive contexts both in and out of class, improving the cognitive and articulatory processes needed for spontaneous speech (Segalowitz, 2010).   This interaction would help participants to build a greater capacity to manage spontaneously constructed online speech – or what we term here "creative competence". However, the greater variability found in development on the unplanned task suggests that creative oral competence remains challenging even after immersion. It also remains an open question how much quality and quantity of exposure can be achieved during the SA experiences of this group of students, and how far this is similar or different to other SA experiences (e.g. Perez Vidal, 2014; Mitchell *et al*., 2017).

In view of the individual variation seen above, we wanted to see how far the participants seemed to have similar degrees of exposure during their months of living in China. It could be that there is a linguistic-affective 'threshold' where those with less confidence or language proficiency found it harder to interact and engage, which then affected their individual rate of development (author and author). We had used a self-report diary protocol (author) aiming to map participants' average levels of interaction over the SA period, along with any qualitative evidence of specific experiences that had affected them. Unfortunately, we were unable to get a full set of these reports; therefore across the group we could not draw any meaningful qualitative themes or quantitative associations between interaction levels and task scores. We found that the highest scoring student at Time 1, reported the highest levels of interaction throughout the period of SA, while the participant with least self-reported interaction had low scores on all tasks, particularly the speaking tasks, at both Time 1 and

Time 2. Further quantitative and qualitative research into the nature of language interaction and experiences in the target country is therefore clearly needed to help clarify the complex interaction of factors that affect individuals' rate of development. In light of the noted cultural and linguistic distance felt by students from US in China (Duff *et al*., 2013; Kinginger, 2011), the challenges facing other western students in China also needs detailed exploration.

Conclusion

The findings reported here, part of a wider study into the development of L2 language for learners of L2 Mandarin, supported the general view that immersion during SA can significantly help learners' oral proficiency development (Collentine & Freed, 2004). We noted task effects played a role too, in that planning time benefited linguistic development, particularly in oral production. This task effect has been noted elsewhere, but not typically examined in an SA context. We suggest that longitudinal SA studies could include task factors such as complexity (Skehan, 2014, author), or comparing monologic vs. dialogic oral fluency (author), given that SA may arguably be the ultimately authentic task-based experience. Comparing oral and written development, we saw that improvements in written tasks seemed to be particularly challenging for UK learners of L2 Mandarin. If written progress in a logographic script seems harder to achieve than oral progress during SA, this is something that institutions may need to consider in their preparation and monitoring of students through their SA. For Chinese SA in particular, the potential effect of differences in writing systems may mean that individual character knowledge and grammatical awareness, apparently secure in untimed and/or planned work may transfer into timed, unplanned writing as commonly assessed in examinations.

We therefore welcome the recent emergence of volumes aiming to create better theory-pedagogy links between L2 Mandarin teaching and SLA-focused research (e.g. Han, 2014; Tao, 2016: Lu, 2017). We call for these links to be further extended to the SA experience, particularly in evaluating the move from instructed classroom to immersion settings aiming to build up cognitive fluency (Segalowitz, 2010) in terms of spontaneous language use or "creative competence". Chinese teachers may retain an attachment to traditional values of drilling and recitation, as seen, for example among Chinese learners of L2 English (Jin & Cortazzi, 2006). We could therefore assume that current expectations of L2 Chinese development may to some extent consist of building up greater skills in producing memorised chunks of language, in speech or writing, where successful reproduction ("performative competence") is evidence of linguistic development. This hypothesis needs testing empirically: if L2 Mandarin language development in pre-SA stages is based on instructed input, students' linguistic proficiency would therefore vary depending on which textbook and what input a teacher would use; this input-based view of learning would predict very different individual outcomes during SA depending on the starting point of linguistic knowledge and practice before arrival. Alternatively, it would be interesting to find evidence suggesting that L2 Mandarin follows a more standardized linguistically-driven developmental route – if so, how far would SA trigger faster development, and how would written or oral mode, or task-based factors have any effects?

The rapid rise in the appeal of Mandarin as a globally significant language provides an excellent good opportunity to bring theory and practice together to clearly assess what constitutes most effective instruction across different global contexts. L2 Mandarin offers thus myriad new, exciting and valuable ways of considering the effects of SA on language learning.

References

Brecht, R., Davidson, D., & Ginsberg, R. (1995). Predictors of foreign language gain during study abroad. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 9-37). Amsterdam: John Benjamins.

Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by compute*r. Version 5.3.51. Retrieved from http://www.praat.org/

Chen, M. (2012). Meiguo liuxuesheng hanyu kouyu chanchu de liuxixing yanjiu [Chinese oral fluency of CSL learners of American English speakers]. *Yuyan Jiaoxue Yu Yanjiu*, (2), 17–24.

Chen, M., & Wang, J. (2008). Hanyu zuowei dier yuyan de kouyu chanchu yunlv bianjie tezheng de gean yanjiu [A case study of prosodic boundary features of spoken Chinese output]. *Hanyu Xuexi*, *4*.

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. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition*, *26*(2), 153–171.

Davidson, D. (2010). Study abroad: When, how long, and with what results? New data from the Russian front. *Foreign Language Annals*, *43*(1), 6–26.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*(01), 5–34.

De Jong, N., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, *61*(2), 533–568.

Du, H. (2013). The development of Chinese fluency during Study Abroad in China. *The Modern Language Journal*, *97*(1), 131–143.

Du, H. (2016). A corpus-approach to the research and teaching of Chinese as a second language. In H. Tao (Ed.), *Integrating Chinese Linguistic Research and Language Teaching and Learning* (pp. 81-96). Amsterdam: John Benjamins.

Duff, P., Anderson, T., Ilnyckyj, R., Van Gaya, E., Wang, R. & Yates, E. (2013). *Learning Chinese: Linguistic, sociocultural, and narrative perspectives*. Berlin/ Boston: DeGruyter.

Ellis, R. (2005). Planning and task-based research: theory and research. In R. Ellis (Ed.), *Planning and task-performance in a second language* (pp. 3-34). Amsterdam: John Benjamins.

Ellis, R., Loewen, S., Elder, C., Philp, J., Reinders, H., & Erlam, R. (2009). *Implicit and explicit knowledge in second language learning, testing and teaching*. Bristol: Multilingual Matters.

Ferris, D. & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, process, and practice.* Mahwah, NJ: Lawrence Erlbaum.

Freed, B., Segalowitz, N., & Dewey, D. (2004). Context of learning and second language fluency in French: Comparing regular classrooms, Study Abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, *26*(02), 275–301.

Hamp-Lyons, L. (Ed.) (1991). *Assessing second language writing in academic contexts*. Norwood NJ: Ablex.

Han, Z. (Ed.) (2014). *Studies in second language acquisition of Chinese*. Multilingual Matters.

Isabelli-García, C. (2010). Acquisition of Spanish gender agreement in two learning contexts: Study abroad and at home. *Foreign Language Annals 43*(2), 289-303.

Jin, L. (2012). When in China, do as the Chinese do? Learning compliment responding in a study abroad program. *Chinese as a Second Language Research 1*(2), 153-316.

Jin, L. & Cortazzi, M. (2006). Changing practices in Chinese cultures of learning. *Language, Culture & Curriculum 19*(1), 5-20.

Kinginger, C. (2011). Enhancing language learning in Study Abroad. *Annual Review of Applied Linguistics*, *31*, 58–73.

Li, W., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. *Neuroscience*, *158*(4), 1416–1425.

Liu, J. (2009). Assessing students' language proficiency: A new model of Study Abroad program in China. *Journal of Studies in International Education*, *14*(5), 528–544.

Lu, Y (Ed.) (2017). *Teaching and learning Chinese in higher education*. Routledge.

MacWhinney, B. (2000). *The CHILDES project: The database* (Vol. 2). Psychology Press.

Mitchell, R., Tracy-Ventura, T., & McManus, K. (2017). *Anglophone Students Abroad: Identity, Social Relationships, and Language Learning*. London: Routledge.

Moyer, A. (2013). An investigation of experience in L2 phonology: Does quality matter more than quantity? *Canadian Modern Language Review*, *67*(2), 191–216.

Myles, F. & Mitchell, R. *French Learner Language Oral Corpora*. http://www.flloc.soton.ac.uk/

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics*, *30*, 590–601.

Perez Vidal, C. (Ed.) (2014). *Language Acquisition in Study Abroad and formal instruction contexts*. Amsterdam: Benjamins.

Sanz, C. (2014). Contributions of Study Abroad research to our understanding of SLA processes and outcomes: The Sala project. In C. Perez Vidal (Ed.), *Language Acquisition in Study Abroad and formal instruction contexts* (pp. 1-13). Amsterdam: Benjamins

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

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

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

Skehan, P. (Ed.) (2014). *Processing perspectives on task performanc*e. London: Benjamins.

Robinson, P. (Ed.) (2001). *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Sun, X. (2008). Zhunbei yinsu dui liuxuesheng hanyu kouyu biaoda de yingxiang [The influence of preparation factors on Chinese oral ability of overseas students]. *Minzu Jiaoyu Yanjiu*, *4*, 89-92.

Tao, H. (Ed.) (2016). *Integrating Chinese linguistic research and language teaching and learning.* Amsterdam: Benjamins.

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

VandeBerg, M., Connor-Linton, J. & Paige, R. (2009). The Georgetown Consortium Project: Interventions for Student Learning Abroad. Frontiers: *The Interdisciplinary Journal of Study Abroad* (special issue), 2-75. http://www.frontiersjournal.com/documents/FrontiersXVIII-Fall09-VandeBerg-ConnorLinton-Paige_000.pdf (accessed 9 May 2013).

Wang, Z. (2014). On-line time pressure manipulations: L2 speaking performance under five types of planning and repetition conditions. In P. Skehan (Ed.), *Processing perspectives on task performance* (pp. 27-61). Amsterdam: John Benjamins.

Xiao, R. & Tao, H. (2006). *The Lancaster Los Angeles Spoken Chinese Corpus*. Lancaster: UCREL, Lancaster University.

Yang, B. (2012). How do pauses reveal linguistic plans by L2 learners of Chinese ? In Q. Ma, H. Ding, & D. Hirst (Eds.), *Proceedings of the 6th International Conference on Speech Prosody* (pp. 390–393). Shanghai: Tongji University Press.

Yang, L. & Sun, M. (2015). Improved learning of Chinese word embeddings with semantic knowledge. In M. Sun, A. Liu, M. Zhang, & Y. Liu (Eds.), *Chinese Computational Linguistics and Natural Language Processing based on Naturally Annotated Big Data.* (pp. 15-25). Dordrecht: Springer.

Yu, B. (2010). Learning Chinese abroad: the role of language attitudes and motivation in the adaptation of international students in China. *Journal of Multilingual and Multicultural Develoment, 31*(3), 301-321.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on Fluency, Complexity and Accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1–27.

Zhao, D., & Liu, Y. (2013). Xiaozu huodong dui shaoshu minzu daxuesheng yingyu kouyu biaoda de yingxiang [The impact of group work on minority college students' oral performance]. *Jincheng Zhiye Jishu Xueyuan Xuebao*, *6*(2).

Effects of Study Abroad on L2 Mandarin Chinese