LANGUAGE LEARNING
*A Journal of Research in Language Studies*

**Language Learning ISSN 0023-8333**

# EMPIRICAL STUDY

# Developing Second Language Mandarin Fluency Through Pedagogic Intervention and Study Abroad: Planning Time, Speech Rate, and Response Duration

Jiayi Wang[a] and Nicola Halenko[b]

[a]De Montfort University [b]University of Central Lancashire

**Abstract:** This longitudinal study examines the effects of a pre-study abroad (SA) pedagogic intervention and subsequent SA experience on second language (L2) Mandarin fluency. It explores two temporal aspects of oral fluency—planning time and speech rate—along with one performance measure, duration of response. Additionally, L2 contact data were included as a supplementary variable in the analysis. The experimental group was assessed at three points: before instruction (T1), after 2 weeks of instruction (T2), and post-SA (T3). A non-instructed control group that participated in the SA period provided baseline data. Both groups demonstrated improved fluency after the SA period, with the experimental group showing superior performance in planning time, speech rate, and duration of response. The greatest reduction in between-group differences occurred at T2 and persisted over time. These findings highlight that combining targeted instruction with exposure is highly effective, with L2 contact strongly correlating with overall fluency gains.

**Keywords** L2 fluency; study abroad; pre-departure instruction; language contact; L2 Chinese/Mandarin

## Introduction

Defined as "a temporary sojourn of pre-defined duration, undertaken for educational purposes" (Kinginger, 2009, p. 11), a period of study abroad (SA)

A one-page Accessible Summary of this article in nontechnical language is freely available in the Supporting Information online and at https://oasis-database.org

Correspondence concerning this article should be addressed to Jiayi Wang, De Montfort University, The Gateway, Leicester, LE1 9BH, United Kingdom. Email: jiayi.wang@dmu.ac.uk

The handling editor for this manuscript was Anne Marie Devlin.





DOI: 10.1111/lang.12694

ostensibly offers language learners daily access to rich, varied, and authentic second language (L2) input. As Sanz (2014) and others have noted, this prime setting aligns with established theories of L2 acquisition as it affords opportunities to notice language (Schmidt, 1990) and negotiate meaning in interaction (Long, 1996), and offers meaningful practice through which linguistic gaps can be uncovered (Swain, 1995). The assumed superiority of this immersive experience for language development has led to continued research interest, particularly as SA research often reports greater variability in linguistic gains than one might expect (Kinginger, 2013; Sanz & Morales-Front, 2018). Variables known to positively influence linguistic gains include the amount of time spent in using the L2 (e.g., Bardovi-Harlig & Bastos, 2011; Du, 2013; Taguchi, Li, & Xiao, 2016), since language practice through exposure and through interaction are known to be effective stimulants of communicative growth (Dekeyser, 2007). Within SA research, speech fluency development is of particular interest due to its complex and multifaceted nature. Listeners engaging with L2 speakers in a SA context may find limited fluency, or displays of disfluency, problematic. Negative reactions may affect potential interlocutors' willingness to interact with L2 learners, which will therefore limit the amount of exposure and input needed to develop L2 fluency in the first place (Derwing, 2017). For all these reasons, speech fluency is an important area of inquiry in SA.

Fluency is often researched via two distinct lenses. A broader understanding of fluency is often linked with notions of global competency or L2 proficiency, as measured by the competency scales of the Common European Framework of Reference for Languages (CEFR), for instance. The narrow view of fluency, as adopted in this study, relates to the ease and fluidity of speech (Lennon, 1990; Segalowitz, 2010). In SA contexts, speech fluency has generally been measured against comparable groups of language learners at home institutions (e.g., Freed, 1995; Freed, Segalowitz, & Dewey, 2004; García-Amaya, 2018; Segalowitz & Freed, 2004) or has been tracked from an acquisitional perspective (e.g., Di Silvio, Diao, & Donovan, 2016; Du, 2013; Freed, Segalowitz, & Dewey, 2004; Mora & Valls-Ferrer, 2012). Differential SA effects are widely reported on a variety of spoken aspects of language (e.g., complexity and accuracy), but investigations of speech fluency at the utterance level appear to consistently report positive gains as a result of short- and long-term immersion in the target language (see García-Amaya, 2018, for a review).

Examining oral fluency is not without its challenges. This study examined one variable impacting oral fluency (planning time), one indicator of oral fluency (speech rate), and one general measure of oral performance (duration). All these examinations can be effectively applied to L2 Mandarin

Chinese, and they collectively offer a holistic indication of time-related and general levels of performance. This paper employs data from a larger project examining how pre-SA training (taken to mean L2 pragmatics instruction received before embarking on a SA period) can benefit learners' linguistic and cultural knowledge of L2 Mandarin. Using data extracted from the larger data set, Wang and Halenko (2022) analyzed the "what" in terms of the quality and content of formulaic sequences from instruction and SA perspectives. In contrast, this study focuses on "how" these sequences were produced in the context of speech fluency through responses to spoken prompts. Given the breadth of SA research that now exists, it is important to expand our knowledge of this specific language learning context to underexplored areas. As noted in García-Amaya's (2018) meta-analysis, most fluency studies use a cross-sectional design, span time periods of less than 6 months in the immersion context, and predominantly focus on Anglophone learners of European languages or on learners of English. The current study addresses all these shortcomings by tracking development of speech fluency longitudinally over a 10-month SA period within the lesser-learned language of Mandarin Chinese. The latter is an important advancement in the field, since China increasingly attracts larger numbers of SA students (Guiaké, Chen, & Zhang, 2021) and holds a strategic place in the global economy, yet only a handful of studies of L2 Mandarin exist (Wang, 2017, 2018; Wang & Halenko, 2019). Another unique feature of the current study is that it combines the influences of a pedagogic intervention and SA on overall speech fluency in spontaneous, asynchronous speech. Wright (2021) recently noted the paucity of research on lexical chunks in Mandarin fluency research, so this paper also makes an important contribution in this area. To the best of our knowledge, few studies have investigated specific links between instruction, fluency, and SA (Tavakoli, Campbell, & McCormack's 2016 paper on L2 learners of English is a notable exception), and none have employed instruction at the pre-departure stage. This is critical as, according to skill acquisition theory, the SA environment greatly facilitates the proceduralization of declarative knowledge acquired during instruction (DeKeyser, 2007). In other words, pre-departure instruction should give L2 users the ability to "hit the ground running" on arrival so they can maximize interaction during SA at the earliest opportunity.

## Background Literature

Much research in speech production derives from Levelt's (1989) model, with processes that move from conceptualization of an idea to conscious formulation, then overt articulation of an utterance. First language (L1) speakers execute these stages simultaneously, at high speed and with minimal effort. This is not the case for L2 speakers, since these processes are still under construction. Movement between these stages can be a stop–start affair, typically requiring more effort and producing language at a slower speed. L2 users also typically have fewer opportunities than L1 speakers to operationalize these processes in real encounters. This section considers fluency in relation to language contact and fluency-related variables within the SA context, before presenting a selected review of fluency studies related to L2 Chinese.

**L2 Fluency and Target Language Contact**

The facilitative but variable effect of target language contact on oral fluency is well documented (e.g., Kinginger, 2013; Mitchell, Tracey-Ventura, & McManus, 2015; Wright, 2021). Variation can be traced to quantity and quality of L2 contact, living situation, and individual characteristics, among other variables (García-Amaya, 2018). A line of studies relevant to the present paper has focused specifically on the effects of "L2 time" on fluency. The amount of language contact or "time-on-task" is increasingly reported to be a more reliable predictor of language success than length of SA stay (e.g., Bardovi-Harlig & Bastos, 2011; Matsumura, 2003; Taguchi et al., 2013, 2016). Language contact data are typically captured via a self-report questionnaire detailing situation-based L2 use, typically the number of hours spent on a variety of activities both inside and outside the classroom (see Arndt, Granfeldt, & Gullberg, 2023, for emerging real-time mobile applications that aim to mitigate against misreporting). The quantitative instrument relies on eliciting learner perceptions of L2 contact; although it risks learners under- or overestimating actual behavior, it has been used successfully in a range of studies of this kind to offer a more detailed picture of *how* learners use their time in the L2 (Bardovi-Harlig and Bastos, 2011; Matsumura, 2003; Taguchi et al., 2013, 2016). Freed, Segalowitz, and Dewey (2004), for instance, found that French L2 learners on an intensive 7-week domestic language immersion program, who regularly accessed the many opportunities to use the L2 inside and outside the classroom, reported more language contact hours and improved the most on fluency variables compared to a group engaged in formal classroom study and a study abroad group. Du (2013), in a study of L2 Chinese participants, concluded that time-on-task was the most important variable in determining fluency development. Du's participants reflected that daily practice allowed them to speak faster and gain confidence in their own abilities to express themselves in the target language.

The benefits of L2 language contact have also been linked to the positive development of formulaic language during SA. Being a member of a target language community can offer easy access to rich and varied formulae through participation in daily communicative events. Additionally, the storage and processing efficiencies of formulaic language for language learners are well known (Kecskes, 2016; Wray, 2002). Matsumura (2003) reported that Japanese learners' knowledge of English advice-giving expressions improved during SA, mediated by their self-reported exposure to English. In findings relevant to the present study, Taguchi, Li, and Xiao (2013) linked learners' improved production of L2 Chinese formulaic expressions for a range of daily communicative encounters (e.g., at a restaurant, at a department store, on a bus) with their perceived frequency of direct contact with the expressions during SA.

In turn, evidence suggests that formulaic language has a positive impact on developing L2 fluency, across a range of discourse features. For instance, Wood (2006) found that the repetition and stringing together of formulaic sequences extended his English learners' lengths of runs between pauses, known to be a key indicator of increased fluency. Yan's (2020) Chinese learners of English produced significantly fewer silent pauses when formulaic sequences were employed, particularly in the construction of longer sentences. Finally, François and Albakry (2021) reported that formulaic sequences were a significant predictor of fluency and allowed learners to incorporate more complex grammar beyond their current level.

## Variables Impacting Fluency in Study Abroad

In quantifying developmental issues with L2 speech, Segalowitz's (2010) conceptualization of fluency is a widely accepted starting point. In this construct, fluency encompasses three distinct but interrelated dimensions: cognitive fluency (the efficient mobilization and integration of the cognitive processes involved in producing utterances), utterance fluency (the objectively measurable aspects of fluency such as speed), and perceived fluency (subjective judgments of L2 speakers' oral fluency). The scope of this paper is primarily concerned with utterance fluency (specifically planning time, speech rate, and duration), since it is a widely adopted approach and can be objectively examined in a robust and systematic way.

### *Planning Time*

Oral fluency has been positively linked with the benefits of planning time for producing higher quality language (Tavakoli & Wright, 2020). In this study, planning time was operationalized on the basis of time taken to mentally prepare an oral response. Levelt (1989) considered the initial stage of planning to be a key component in the complex process of language production, as it allows language users time to scan and select the appropriate linguistic conventions to express their communicative intent. Pre-task planning is specifically known to benefit fluency (Tavakoli & Skehan, 2005) and can be particularly useful at lower proficiency levels. In Ellis's (2009) review of 19 studies examining the effects of offering learners pre-task planning, 17 studies showed that planning contributed to increased fluency. Linking SA, planning time, and proficiency with the production of requests in L2 Chinese, Li (2014), however, reported that a 15-week SA program for intermediate and advanced L2 Chinese learners had no effect on reducing planning time, possibly due to anxiety or personality variables.

*Speech Rate*

The second aspect of fluency employed in this study is speech rate, calculated as the number of syllables or words uttered per minute or second. Employed widely in fluency studies (Derwing, 2017; Tavakoli & Wright, 2020), speech rate has previously been found to be one of the most important variables perceived by L1 users for determining fluent speech (Freed, Segalowitz, & Dewey, 2004). Although Skehan (2014) has proposed adopting utterance fluency measures that combine both speed and flow (breakdown and repair measures), these were not considered viable because the study design utilized a computer-mediated elicitation task, which is known to capture responses containing less repair and breakdown than face-to-face interaction (Tavakoli & Wright, 2020). The composite measure of speech rate that combines pausing and speed aspects of fluency (de Jong et al., 2012) was, however, appropriate for the study's aims. Fluency studies incorporating measures of L2 speech rate have overwhelmingly reported significant increases as a result of SA across a range of languages and SA lengths (e.g., Du, 2013; Huensch & Tracy-Ventura, 2017; Mora & Valls-Ferrer, 2012).

*Duration*

The final fluency variable employed was duration of speech, which captures the total time taken to produce speech including all hesitation phenomena. Duration was calculated to represent overall performance of the learner groups. Following Freed, Segalowitz, and Dewey (2004), duration and speech rate were recorded alongside one another to capture both a general indication of oral performance (duration) and a time-related specific measure of oral speed (speech rate) for a more holistic examination of instruction and SA effects. Mora and Valls-Ferrer (2012) showed that advanced-level Catalan learners of English made robust gains in oral fluency in time-related aspects of speech production, including speech rate and duration, compared to an at-home group receiving instruction. Freed, Segalowitz, and Dewey's (2004) findings showed that, compared to a group engaged in formal classroom study and a study abroad group, the domestic language immersion group made the most gains across a number of variables, including speech rate.

### L2 Chinese Fluency

Developing fluency in a L2 that is typologically distant from the L1, such as L2 Mandarin Chinese in relation to L1 English in this study, poses an additional burden on the learning process (Derwing, Munro, Thomson, & Rossiter, 2009) and affects learners at all stages of Levelt's (1989) speech production construct, as described earlier. Tavakoli and Wright (2020) highlighted specific issues in L2 Mandarin with word order rules requiring whole utterance planning prior to articulation, high levels of polysemy making sound–meaning pairings challenging, and a tone system requiring accurate articulation of words to distinguish meanings. These additional challenges have the potential to limit learners' confidence and ability to interact during SA. L2 Mandarin research, as addressed in the following studies, is therefore of high value in showing how such challenges unfold in real time.

Kim et al. (2015) analyzed several aspects of L2 Chinese development, including a range of fluency variables, before and after a semester-long SA period. Simulated oral proficiency interviews captured data from 22 L2 learners of Chinese from a U.S. university's SA program in China. Findings highlighted that speech rate was a significant predictor of fluency gains, unlike the other temporal aspects of production (filled/unfilled pauses and mean pause length), which yielded mixed results. Du's (2013) examination of speech rate further highlighted that L2 Chinese interaction was the most important variable in determining fluency development. Interestingly, the greatest speech rate gains were noted in the 1st month of SA, which the author suggested was a period of settlement into the new environment, where frequent and novel communicative challenges forced more regular L2 output, leading to greater fluency.

A more recent study by Wright (2021), with a participant group of adult L2 Chinese learners from a British university on a 10-month SA, shares similarities with the present paper. Her pre–post investigation, however, focused on the impact of SA on oral tasks (rehearsed vs. spontaneous and dialogic vs. monologic) and different aspects of utterance and breakdown fluency variables. Overall, despite performance in rehearsed tasks being generally superior to spontaneous speech pre- and post-SA, immersion seemed to clearly aid fluency in terms of spontaneous speech. Wright concluded that SA appeared to foster greater creative communicative competence in learners' spoken output.

Since language learners appear to develop oral fluency as a result of SA, it would be useful to know to what extent pre-SA instruction contributes to increased fluency development beyond any SA gains achieved without this intervention. Pre-SA instruction is reportedly effective for developing pragmatic competence in general (Wang & Halenko, 2019, 2022), but, to date, this has yet to be explored as a variable impacting fluency. Naturally, language learners who have received instruction beforehand might typically hold an advantage over a non-instructed group, certainly in the initial stages of SA. But is simple exposure to the L2 and immersion in the target language environment sufficient to match any fluency gains held because of prior intervention? This was the focus of the present study, which was guided by the following research questions:

1. To what extent is planning time in L2 Mandarin influenced by SA and pre-SA instruction?
2. To what extent is speech rate in L2 Mandarin influenced by SA and pre-SA instruction?
3. To what extent is duration of response in L2 Mandarin influenced by SA and pre-SA instruction?
4. To what extent does the amount of contact with the L2 impact fluency variables?

## Method

### Participants

The participants were undergraduate students of L2 Mandarin Chinese at a British university. At the time of the study, the 18 students (9 males and 9 females), aged 21–32 years, were studying Chinese as a L2 as part of their degree programs. The sample included 15 British students who were native English speakers and three European students (French, German, and Italian) who had been living and studying in the UK for over 10 years. All participants had completed 2 years of L2 Chinese study, achieving an average upper-intermediate proficiency level. None of the students had prior SA experience in China exceeding 2 weeks, nor had they participated in any such experiences within the 12 months preceding the study.

The students were assigned to either an experimental group (EG; $n = 9$), which received a pre-SA instructional intervention, or a control group (CG; $n = 9$), which did not receive any instructional treatment. This assignment followed methods similar to those used by Li, Ellis, and Kim (2018) and by

Nergis (2021) and ensured an even distribution of Chinese achievement levels and group homogeneity in terms of performance.[1]

**Pre-Study-Abroad Instruction and Oral Tests**

Chinese textbooks do not often pay special attention to formulaic expressions, but "Chinese language learners do, especially if they study abroad" (Kecskes, 2016, p. 117). There was repeated feedback from year-abroad returners that they did not feel adequately equipped to say the appropriate Chinese phrases needed to do everyday things in China, and it was this that motivated us to conduct pre-SA instruction. This study aimed to better prepare students to produce formulaic speech for their SA.

The pre-SA sessions involving explicit pragmatic instruction were spread over 2 weeks (6 hours in total) and were co-delivered by a native speaker of Chinese and a native speaker of English. The sessions were organized according to seven communicative themes that captured 26 different transactional and social expressions: (a) compliment response, (b) request, (c) inquiry, (d) leave-taking, (e) telephone conversation, (f) bargaining, and (g) apology. Metapragmatic knowledge, covering speech act strategy selection and the linguistic means to realize these, was also taught (see Wang & Halenko, 2019, for details).

The formulaic expressions used in this study were developed through consultation with the literature, SA returners, and native and non-native speakers of Chinese. Following Taguchi et al. (2013) and Bardovi-Harlig (2009), 26 situations and their target phrases were selected after piloting (see Wang & Halenko, 2022, for details), as illustrated by the following example:

> *Scenario 3: In a restaurant*
>
> You are having dinner in a restaurant. You ask the waitress to take the leftovers with you. You say?
>
> Target phrase: 打包 (Wrap up)

For the testing phases, the participants completed a computerized oral test, and responses were recorded using Audacity software. The test involved reading, listening, and responding to a series of day-to-day scenarios (Appendix S1 in the Supporting Information online). Each scenario contained a description with a subject line, a paragraph describing the scenario, and a picture representing the scenario. Participants were asked to imagine themselves in these situations in China and produce an appropriate oral response in Chinese when they heard the prompt "You say?". Learners worked through a PowerPoint presentation illustrating the 26 scenarios at their own pace and took 17 min on average to complete the test. The test was administered three times with both the EG, which received the instruction, and the CG, which did not: before the instruction (Time 1), immediately after the 2-week pre-SA instruction (Time 2), and following the completion of their SA, 1 year later (Time 3).

The computerized oral test elicited a single-turn oral response to each scenario. Since the current study was not examining interactional or collaborative features of talk between speakers, this task allowed examination of output without interruptions from an interlocutor, thereby also offering a higher degree of control and predictability of outcomes to compare between-learner output more effectively and efficiently (de Jong et al., 2012). This task type met the requirements of the study, was more suitable for a computer-based elicitation task, and avoided stretching the learners' cognitive load while on-task. Although spontaneous speech may increase cognitive load, attempts were made to mitigate this by providing clearly structured, short, repetitive tasks so that attention could be allocated more directly to language formulation and production (Tavakoli & Skehan, 2005).

The computerized oral test contained the same scenarios in each test stage, but the order of the scenarios was changed each time to mitigate any test effects. This study advanced the use of computerized oral tests to capture data for fluency studies and thus makes a methodological contribution, since most L2 fluency studies opt for oral proficiency interviews, simulations of these, or monologic data collection methods such as picture narratives to elicit oral data.

To ensure the reliability of the instrument, an assessment was conducted, which demonstrated high reliability with a Cronbach's alpha of .844 for the 26 items. This indicates that the instrument is consistent and reliable for measuring the constructs of interest.

## L2 Fluency Variables

The participants' oral output was examined using three metrics of L2 fluency: two time-related aspects of oral fluency (planning time and speech rate) and one general measure of oral performance (duration). First, planning time referred to the number of seconds taken to prepare for each scenario and before uttering the first syllable. In Chinese, a syllable (which represents a character) can be a word on its own. It can also be combined with one or more other syllables to form a word. Second, speech rate was calculated by dividing the number of syllables produced by utterance duration. The mean speech rate referred to the average number of syllables uttered per minute. Third, duration of response referred to the time span from the start of an utterance to the end, including pauses and hesitations, but excluding the pre-task planning time. Duration metrics are reported in seconds.

The L2 fluency data are expressed as means and standard deviations. The small sample size of this study did not allow for a multivariate longitudinal analysis. A repeated measures analysis of variance (ANOVA) was therefore used with two groups (EG and CG) and three sample times (T1: baseline before the pre-SA instruction; T2: immediately after the 2-week instruction; T3: after the year abroad). The data were normally distributed (see Appendix S2 in the Supporting Information online for skewness, kurtosis, and $z$-score data). A $p$ value of $< .05$ was considered significant.

### Language Contact

Finally, the language contact profile (Appendix S3 in the Supporting Information online) elicited information about both in-class and out-of-class Chinese language contact as well as information about English language contact time. The survey employed here is an adapted version of Freed et al.'s (2004) survey. Specifically, we omitted sections about accommodation, including home-stay and apartment types with native or non-native speakers, because all participants stayed in university-allocated apartments for foreign students. Learners were required to estimate and record the number of hours per day and days per week for which the target language was used. The amount of L2 contact per week during SA for each participant was calculated by taking the sum of all the different types of Chinese language contact reported, including the time spent communicating in Chinese with classmates, friends, teachers, and service personnel. The survey was administered twice: before the pre-SA instruction (T1) and after SA (T3).

## Results

This study analyzed the impact of pre-SA instruction and SA on planning time, speech rate, and duration of response as aspects of L2 Mandarin fluency. To help interpret gains in oral fluency and performance, we also correlated the data with the amount of time spent using L2 Mandarin during SA.

To provide an initial overview of the main results, Table 1 summarizes the descriptive statistics of the three fluency measures for the responses from the two participant groups. Following this, the results for planning time, speech rate, duration, and L2 contact are presented in turn.

### Planning Time

Both the EG and the CG reduced their pre-task planning time throughout the study period: from before instruction (T1) to immediately after 2 weeks'

**Table 1** Descriptive statistics: Second language (L2) fluency measures for the experimental and control groups at Times 1, 2, and 3

| | | Experimental group | | | Control group | | |
|---|---|---|---|---|---|---|---|
| L2 fluency measures | Time | *M* | *SD* | 95% CI | *M* | *SD* | 95% CI |
| Planning time | T1 | 4.95 | 3.96 | [3.13, 6.77] | 5.07 | 2.65 | [4.10, 6.04] |
| | T2 | 1.31 | 0.37 | [1.24, 1.38] | 3.42 | 1.80 | [2.75, 4.09] |
| | T3 | 1.09 | 0.20 | [1.05, 1.13] | 1.92 | 0.76 | [1.53, 2.31] |
| Speech rate | T1 | 132.81 | 21.71 | [121.62, 143.99] | 105.86 | 52.21 | [84.04, 127.68] |
| | T2 | 211.89 | 38.68 | [193.55, 230.23] | 120.15 | 37.44 | [106.47, 133.83] |
| | T3 | 222.68 | 21.47 | [213.52, 231.83] | 160.82 | 51.01 | [135.60, 186.03] |
| Duration | T1 | 173.53 | 62.08 | [147.27, 199.80] | 302.53 | 322.90 | [184.71, 420.36] |
| | T2 | 43.58 | 9.02 | [40.16, 47.00] | 223.59 | 275.08 | [103.90, 343.27] |
| | T3 | 68.71 | 16.55 | [62.86, 74.56] | 117.76 | 43.57 | [107.36, 128.15] |

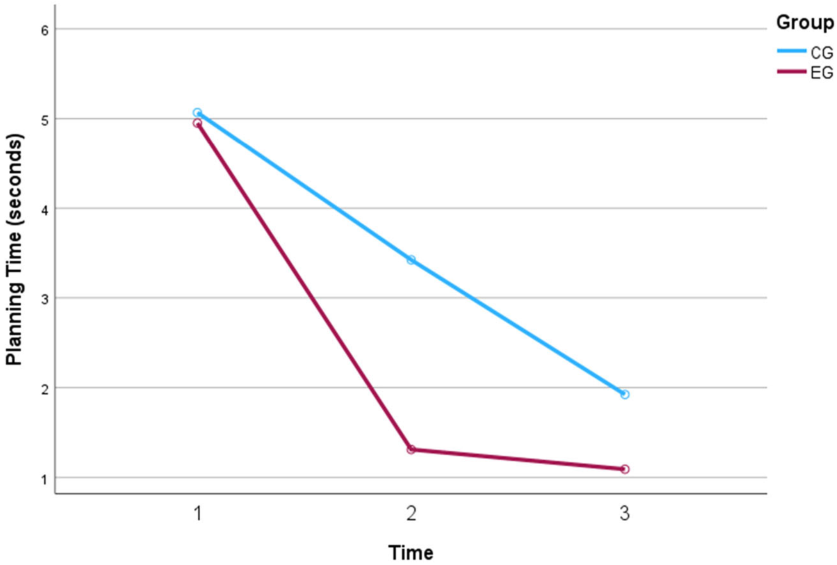


**Figure 1** Estimated marginal means of planning time (in seconds). CG = control group; EG = experimental group.

instruction (T2), and then to after the year abroad (T3). Both groups were able to execute utterances more efficiently over time, with greater within-group variability reduction for the EG than for the CG, as indicated by the standard deviations. The level of improvement differed noticeably between the two groups, as Figure 1 illustrates.

As Figure 1 shows, the EG decreased its pre-task planning time more sharply than the CG did post-instruction, but the improvement slowed down during the year abroad. The planning time data were submitted to a repeated measures ANOVA. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(2) = 0.249$, $p < .001$, and therefore a Greenhouse–Geisser correction was used because $\varepsilon < .75$. There was a significant effect of time on pre-task planning time with a large effect size, $F(1.140, 18.241) = 17.632$, $p < .001$, $\eta_p{}^2 = .524$. This implies that the effects of pre-SA instruction and SA on the mean planning time were large.

Pairwise comparisons, with Bonferroni adjustments made for multiple comparisons, revealed that the EG made significant improvements in planning time between T1 ($M = 4.95$, $SD = 3.96$) and T2 ($M = 1.31$, $SD = 0.37$), with a mean difference of 3.64, 95% CI [2.28, 5.00], $t(16) = 8.215$, $p < .001$, Cohen's $d = 1.62$; and between T1 and T3 ($M = 1.09$, $SD = 0.20$), with a mean difference of 3.86, 95% CI [2.44, 5.28], $t(16) = 7.846$, $p < .001$, Cohen's $d = 1.72$. However, there was no significant change between T2 and T3, with a mean difference of 0.22, 95% CI [−0.34, 0.78], $t(16) = 1.357$, $p = .367$, Cohen's $d = 0.42$. This indicates significant improvements post-instruction and throughout the whole study period for the EG, but not during the SA period.

In contrast, the CG showed no significant improvement in planning time between T1 ($M = 5.07$, $SD = 2.65$) and T2 ($M = 3.42$, $SD = 1.80$), with a mean difference of 1.65, 95% CI [−0.51, 3.81], $t(16) = 1.356$, $p = .188$, Cohen's $d = 0.68$. However, significant improvements were observed between T2 and T3 ($M = 1.92$, $SD = 0.76$), with a mean difference of 1.50, 95% CI [0.50, 2.50], $t(16) = 5.318$, $p < .001$, Cohen's $d = 0.89$; and between T1 and T3, with a mean difference of 3.14, 95% CI [0.88, 5.40], $t(16) = 4.818$, $p = .003$, Cohen's $d = 1.18$.

The effect sizes, indicated by partial eta squared, were substantial for the EG between T1 and T2 ($\eta_p^2 = .572$, 95% CI [0.30, 0.70]) and between T1 and T3 ($\eta_p^2 = .524$, 95% CI [0.24, 0.66]), whereas the CG showed significant effects between T2 and T3 ($\eta_p^2 = .268$, 95% CI [0.05, 0.52]) and between T1 and T3 ($\eta_p^2 = .391$, 95% CI [0.10, 0.58]). The EG benefited significantly from both the pre-SA instruction and the entire study period, whereas the CG benefited significantly from the SA period.

Both groups were similar at the beginning of the study, as no statistically significant main effect of group was found for T1, $F(1, 16) = 2.148$, $p = .162$, $\eta_p^2 = .118$, 95% CI [0.00, 0.33]. The main effect of group became significant with a large effect size for T2, $F(1, 16) = 11.939$, $p = .003$, $\eta_p^2 = .427$, 95% CI [0.20, 0.62], and T3, $F(1, 16) = 10.215$, $p = .006$, $\eta_p^2 = .390$, 95% CI [0.15, 0.58]. At the posttest stage after the instruction (T2), the EG was significantly faster than the CG in planning time, indicating the effects of the pre-SA instruction. At the delayed posttest stage after SA (T3), the gap between the two groups narrowed, but the EG's planning time remained significantly shorter than that of the CG. This suggests that the CG's mean planning time caught up with that of the EG to some extent, facilitated by the SA, and that the EG did not gain as significantly from SA as the CG did. Arguably, the higher the level, the slower the overall rate and degree of progression.

### Speech Rate

The speech rate results showed that both groups improved at all stages, but to different degrees. Figure 2 depicts the trends of the estimated marginal means of speech rate increases over time. The EG's increase slowed during the year abroad, whereas the CG's growth accelerated during the same period.

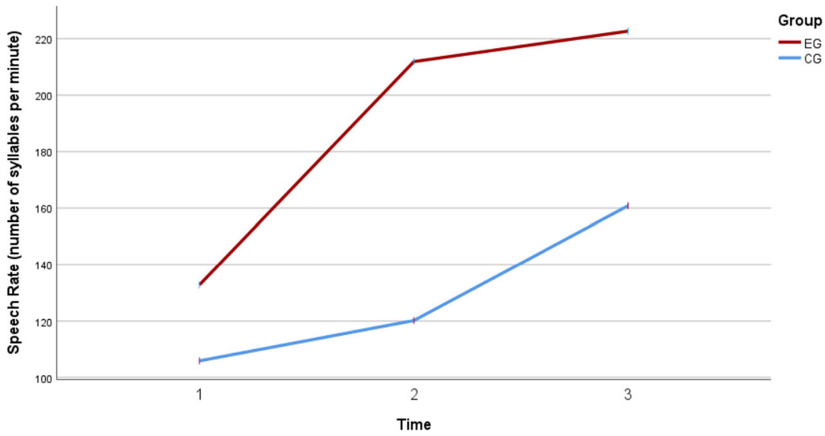


**Figure 2** Estimated marginal means of speech rate (number of syllables per minute). EG = experimental group; CG = control group.

The speech rate data were submitted to a repeated measures ANOVA. Mauchly's test of sphericity indicated that the assumption of sphericity was met, $\chi^2(2) = 2.494$, $p = .287$. The main effect of time yielded an effect size of .678, 95% CI [0.42, 0.80], implying that 67.8% of the variance in mean speech rates was explained by time, $F(2, 32) = 33.721$, $p < .001$, $\eta_p{}^2 = .678$. The main effect of group yielded an effect size of .493, 95% CI [0.19, 0.69], indicating that 49.3% of the variance in the mean speech rates was explained by group, $F(1, 16) = 15.569$, $p = .001$, $\eta_p{}^2 = .493$. The interaction effect was significant, $F(2, 32) = 6.577$, $p = .004$, $\eta_p{}^2 = .291$, 95% CI [0.07, 0.55], implying that there was a significant combined effect for group and time on the speech rate. This implies that the effects of pre-SA instruction and SA on the mean speech rates were substantial.

Pairwise comparisons with Bonferroni adjustments for multiple comparisons reveal that the EG made significant improvements in speech rate between T1 ($M = 132.81$, $SD = 21.71$) and T2 ($M = 211.89$, $SD = 38.68$), with a mean difference of 79.08, 95% CI [56.50, 101.66], $t(16) = 8.509$, $p < .001$, Cohen's $d = 2.12$; and between T1 and T3 ($M = 222.68$, $SD = 21.47$), with a mean difference of 89.88, 95% CI [66.58, 113.17], $t(16) = 9.615$, $p < .001$, Cohen's $d = 2.33$. However, there was no significant change between T2 and T3, with a mean difference of 10.79, 95% CI [−12.52, 34.10], $t(16) = 1.215$, $p = .239$, Cohen's $d = 0.30$. This indicates significant improvements post-instruction and post-SA for the EG, but not during the SA period.

In contrast, the CG showed no significant improvement in speech rate between T1 ($M$ = 105.86, $SD$ = 52.21) and T2 ($M$ = 120.15, $SD$ = 37.44), with a mean difference of 14.30, 95% CI [−16.43, 45.02], $t(16) = 0.657$, $p = .524$, Cohen's $d = 0.30$. However, significant gains were observed between T1 and T3 ($M$ = 160.82, $SD$ = 51.01), with a mean difference of 54.96, 95% CI [18.53, 91.39], $t(16) = 3.257$, $p = .005$, Cohen's $d = 1.08$; and between T2 and T3, with a mean difference of 40.67, 95% CI [6.81, 74.54], $t(16) = 2.577$, $p = .020$, Cohen's $d = 0.87$.

The effect sizes, indicated by partial eta squared, were large for the EG between T1 and T2 ($\eta_p^2 = .678$, 95% CI [0.42, 0.80]) and between T1 and T3 ($\eta_p^2 = .693$, 95% CI [0.46, 0.82]), whereas the CG showed smaller but significant effects between T1 and T3 ($\eta_p^2 = .422$, 95% CI [0.11, 0.63]) and between T2 and T3 ($\eta_p^2 = .294$, 95% CI [0.02, 0.52]). The EG gained the most from the instruction, whereas the CG benefited the most from the SA.

Both groups were comparable at the beginning of the study, and no statistically significant main effect of group was found for T1, $F(1, 16) = 2.044$, $p = .172$, $\eta_p^2 = .113$, 95% CI [0.00, 0.35]. However, the main effect of group became significant with a large effect size for T2, $F(1, 16) = 26.129$, $p < .001$, $\eta_p^2 = .62$, 95% CI [0.33, 0.77], and T3, $F(1, 16) = 11.245$, $p = .004$, $\eta_p^2 = .413$, 95% CI [0.12, 0.64]. At the posttest stage post-instruction (T2), the EG was significantly faster than the CG, implying the effects of the pre-SA instruction. At the delayed posttest stage after SA (T3), the gap between the two groups narrowed, but the EG produced significantly more syllables per minute than the CG did. Perhaps the pre-SA instruction enabled the EG to maintain their competitive edge to an extent even after SA, but it also made it harder for the EG to further improve their speech rate significantly during SA.

### Duration

The duration of response results showed that the EG's duration was the shortest at the posttest stage (T2). However, the EG reversed this reduction during the SA period, leading to an increase at T3. In contrast, the CG's duration was the shortest at the final stage (T3), showing a continuous reduction throughout the study period, as shown in Figure 3.

The duration data were submitted to a repeated measures ANOVA. Mauchly's test of sphericity indicated that the assumption of sphericity had been violated, $\chi^2(2) = 0.306$, $p < .001$, and therefore a Greenhouse–Geisser correction was used as $\varepsilon < .75$. There was a significant effect of time on duration, with a large effect size, $F(1.180, 18.885) = 6.930$, $p = .013$, $\eta_p^2 = .302$,

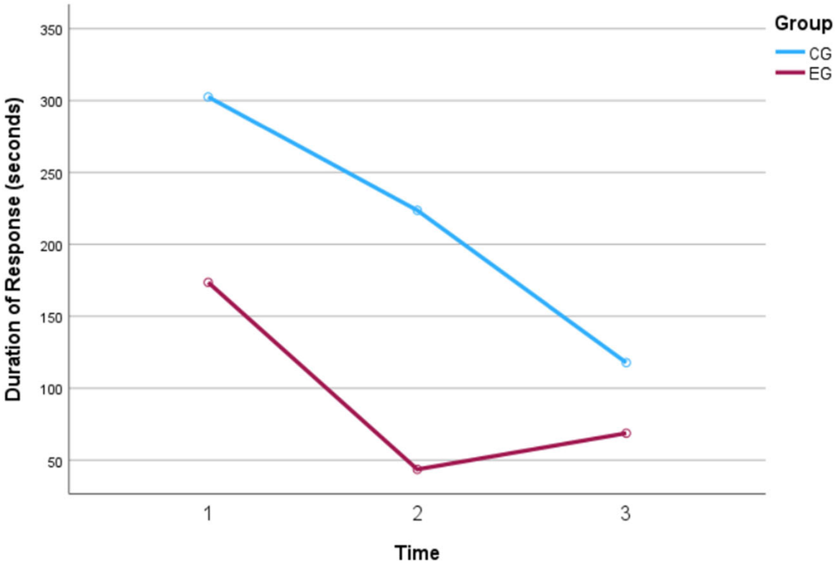


**Figure 3** Estimated marginal means of duration of response (seconds). CG = control group; EG = experimental group.

95% CI [0.07, 0.53]. This implies that the effects of pre-SA instruction and SA on the duration of the response were large.

Pairwise comparisons, with Bonferroni adjustments made for multiple comparisons, revealed that the EG significantly shortened their response duration between T1 ($M = 173.53$, $SD = 62.08$) and T2 ($M = 43.58$, $SD = 9.02$), with a mean difference of 129.95, 95% CI [88.60, 171.29], $t(16) = 7.784$, $p < .001$, Cohen's $d = 2.10$; and between T1 and T3 ($M = 68.71$, $SD = 16.55$), with a mean difference of 104.82, 95% CI [4.12, 205.52], $t(16) = 2.586$, $p = .047$, Cohen's $d = 1.18$. However, there was no significant change between T2 and T3, with a mean difference of 25.13, 95% CI [−50.87, 101.13], $t(16) = 0.684$, $p = .504$, Cohen's $d = 0.58$. This indicates significant improvements post-instruction and throughout the whole study period for the EG, but not during the SA period, where a reversal in trend was observed.

In contrast, the CG showed no significant improvement in response duration between T1 ($M = 302.53$, $SD = 322.90$) and T2 ($M = 223.59$, $SD = 275.08$), with a mean difference of 78.94, 95% CI [−5.19, 163.07], $t(16) = 1.352$, $p = .193$, Cohen's $d = 0.34$. However, significant reductions were observed between T2 and T3 ($M = 117.76$, $SD = 43.57$), with a mean difference of 105.83, 95% CI [12.53, 199.13], $t(16) = 2.918$, $p = .010$, Cohen's $d = 0.79$, and throughout the whole study period between T1 and T3, with a mean difference of 184.78, 95% CI [15.69, 353.87], $t(16) = 3.873$, $p = .002$, Cohen's $d = 1.15$.

The effect sizes, indicated by partial eta squared, were substantial for the EG between T1 and T2 ($\eta_p^2 = .749$, 95% CI [0.50, 0.85]) and between T1 and T3 ($\eta_p^2 = .455$, 95% CI [0.15, 0.68), whereas the CG showed significant effects between T2 and T3 ($\eta_p^2 = .358$, 95% CI [0.06, 0.59]) and between T1 and T3 ($\eta_p^2 = .484$, 95% CI [0.10, 0.70]). The EG benefited significantly from both the pre-SA instruction and the entire study period, whereas the CG benefited significantly from the SA period.

Both groups were similar at the beginning of the study, and no statistically significant main effect of group was found for T1, $F(1, 16) = 1.385$, $p = .256$, $\eta_p^2 = .080$, 95% CI [0.00, 0.30]. The main effect of group became significant with a large effect size for T2, $F(1, 16) = 3.850$, $p = .046$, $\eta_p^2 = .194$, 95% CI [0.00, 0.45], and T3, $F(1, 16) = 9.965$, $p = .006$, $\eta_p^2 = .384$, 95% CI [0.08, 0.61]. At the posttest stage after the instruction (T2), the EG's duration of response was significantly shorter than that of the CG, indicating the effects of the pre-SA instruction. At the delayed posttest stage after SA (T3), the gap between the two groups narrowed, but the EG's duration of response was still shorter than that of the CG. The CG had caught up with the EG to some extent, facilitated by the SA.

Interestingly, the reduction in the EG's duration backtracked between T2 and T3, whereas even without the pre-departure intervention, the CG shortened its duration in the same period. The EG peaked at the posttest stage immediately after the instruction (T2). However, it is important to consider the qualitative results in addition to the quantitative measures. This is explored further in the Discussion section. It is noteworthy that the EG, despite its move backwards during the SA, still outperformed the CG post-SA (T3) in terms of duration of speech.

In short, the pre-SA instruction had a significant impact on the EG's planning time, speech rate, and duration of response, but the SA did not facilitate the same levels of positive impact on the EG as it did for the CG. In fact, the SA even reversed the improvement of the EG's duration, the reasons for which are further explored in the Discussion. Nonetheless, throughout the entire study period between T1 and T3, the EG showed significant development in all three areas, demonstrating the usefulness of pre-SA instruction. The CG demonstrated significant improvement in all three areas through the SA, highlighting the significant impact of SA on the three measures of fluency, even without

**Table 2** Reported second language contact (hours per week)

| Time | Group | *M* | *SD* | Min. | Max. | Mean 95% CI |
|---|---|---|---|---|---|---|
| T2 | EG | 31.11 | 17.97 | 7 | 55 | [22.17, 40.05] |
| | CG | 28.22 | 10.33 | 10 | 48 | [23.08, 33.36] |
| T3 | EG | 35.56 | 16.99 | 11 | 57 | [27.11, 44.01] |
| | CG | 45.56 | 28.52 | 15 | 107 | [31.38, 59.74] |

*Note*. $N = 18$. EG = experimental group; CG = control group.

pre-SA instruction. However, the EG remained superior to the CG throughout the study period, demonstrating the usefulness of combining pre-SA instruction with SA.

## Language Contact

The total amount of L2 contact was not significantly different between the EG and the CG either before or during the SA, as confirmed by the results of the independent *t* tests before SA (T2), $t(16) = -0.418$, $p = .681$, Cohen's $d = 0.197$, 95% CI [−15.97, 12.66], and after SA (T3), $t(16) = -0.904$, $p = .380$, Cohen's $d = 0.426$, 95% CI [−31.23, 12.34]. The two participant groups were similar in terms of their L2 contact during the period abroad. To further analyze the data, we examined the extent to which the positive gains in planning time, speech rate, and duration were correlated with L2 contact during the SA. The Pearson correlation results revealed that the amount of L2 contact during SA correlated strongly with the gains in speech rate ($r = .66$, $p = .003$), but not with the improvements in planning time ($r = .02$, $p = .950$) or changes in duration ($r = -.18$, $p = .490$). It is worth pointing out that during SA, that is, between T2 and T3, the individual differences grew larger for the CG but not for the EG, as shown by the maximum hours reported and the standard deviations in Table 2 on L2 contact.

As shown in Table 2, the maximum reported L2 contact during SA was 107 hours per week, which was substantially higher than the average contact of 11 to 64 hours per week for other participants. Notably, the CG participant with the highest L2 contact, who majored in law with Chinese and was the weakest student before SA, committed a lot of time to writing essays or homework tasks in Chinese outside class (e.g., "until it was complete and until I felt I could understand it"). Despite lacking pre-SA instruction, she achieved substantial improvements post-SA, with the second-largest improvement in speech rate and the fourth-largest improvement in planning time. Her individual improvements were strongly correlated with her L2 contact.

A linear regression analysis was performed to assess the impact of L2 contact on the gain in speech rate during SA. The model accounted for 63.6% of the variance in speech rate gains ($R^2 = .636$, adjusted $R^2 = .601$) and was statistically significant, $F(1, 16) = 12.383$, $p = .003$. The partial eta squared for L2 contact was .603, 95% CI [0.35, 0.78], indicating a substantial effect size. The coefficient for L2 contact was 1.158 ($SE = 0.329$), with a 95% confidence interval of [0.461, 1.855], and was a significant predictor, $t(16) = 3.519$, $p = .003$, suggesting that increased L2 contact is associated with greater improvements in the average speech rate between T2 and T3. The intercept of the model was −21.234, suggesting a baseline loss or very low gain in speech rate when L2 contact is zero, although this value was not statistically significant ($p = .184$).

Overall, across both groups, L2 contact during SA was strongly correlated with gains in speech rate but not with changes in planning time or duration. The observed individual differences underscore the need for further investigation.

## Discussion

This study sought to quantify growth in L2 Mandarin fluency as a result of a pedagogic intervention and SA experience. These findings were compared to those for participants who only experienced immersion in the SA context, without the benefit of a prior intervention. Planning time, speech rate, and duration were used as fluency-related variables to evidence change and are discussed in turn in the following sections. The impact of language contact is also examined.

### Planning Time

While still based in the UK (T1–T2), both the experimental group and control group evidenced a reduction in pre-task planning time. For the control group, this change was not significant. Test effects may be the likely explanation for the control group's reduction given the short time lapse between test phases, in addition to task repetition, which is also known to free up attentional resources (Bygate, 2001). Although this may also partially explain the experimental group's significant reduction in planning, the significant T1–T2 and T2–T3 between-group differences suggest additional influential variables. Turning to existing SA research, neither Li (2014) nor Taguchi (2007) found any significant SA effects on pre-task planning time with either L2 Chinese or L2 English participants, respectively, as a result of L2 exposure and interaction alone. In both studies learners generally maximized the unlimited time provided to formulate their responses. It might follow then that the instructional input and/or the processing and retrieval advantages of formulaic language chunks (Pawley & Syder, 1983; Wray, 2002) could account for lower usage of planning time found in both of the group's data.

The pre-SA trends in reducing planning time continued for both groups post-SA (T2–T3). Given the 10-month time lapse, however, the SA setting may have been the primary contributor this time, particularly for the control group, whose planning time was at the lowest point on return from SA. We can interpret these findings in several ways, but can rule out language contact as an indicator, since no correlations between L2 contact and planning time were found. First, the T3 language contact questionnaires also captured data on how frequently the students encountered the day-to-day scenarios during SA, as they appeared on the oral test. As findings show, the more frequently the participants encountered a scenario during SA, the more likely they performed better in their planning time for that scenario after SA. It is possible the awareness-raising effect of taking the test twice before SA primed the control group to notice the kinds of necessary functional language needed during SA, which led to attending to their linguistic gaps while in situ. Another plausible explanation is that, similarly to Du's (2013) students, the control group grew in confidence and willingness to take risks over time. Comments from the questionnaires include "I was really scared that nobody would understand my Chinese, but the year abroad helped me to overcome my fears. I became more and more confident to speak Chinese" and "I had to handle everything myself. … I grew more and more willing to take risks." Positive psychological effects of SA have been reported elsewhere, with participants feeling more confident, resourceful, and autonomous on return from SA (Kinginger, 2013; Mitchell, Tracy-Ventura, & McManus, 2015).

T1–T3 findings still show that the experimental group significantly outperformed the control group on planning time. Despite both groups becoming more efficient in knowledge retrieval, the quality of oral output showed a marked difference between the groups. Analyzing the formulaic output of these same groups, Wang and Halenko (2022) noted that the experimental group tended to produce responses that were more sophisticated and elaborate than those of the control group. For instance, as previously analyzed by Wang and Halenko (2022), an experimental group member's response to Scenario 18, Bargaining ("In a market, you want to buy a T-shirt but you think it's a bit expensive. You want the vendor to lower the price"), was "老板我是学生, 便宜点儿嘛" ("Boss I am a student, cheap a bit AUXILIARY"). This response was richer than the taught target phrase, incorporating terms of address (e.g., 老板 "boss") and reasons/justifications (e.g., 我是学生 "I'm a student"), which were not explicitly taught in the pre-SA training. The CG findings are reminiscent of so-called "trade-off" effects where aspects of fluency, such as planning time, are attended to at the expense of the complexity and accuracy of the utterance (Yuan & Ellis, 2003).

**Speech Rate**

The EG data suggest they were consistently more fluent in terms of speech rate, though this was a mix of significant and nonsignificant differences. At T1–T2, the experimental group produced faster responses compared to the CG, but nonsignificant differences were found within both groups (EG: T2–T3; CG: T1–T2). At both T2 and T3, the experimental group outperformed the control group, suggesting they profited from the added value of instruction. As formulaic language is known to extend the capacity for fluent production (Wray, 2002), the pre-SA training may have triggered more automatic behavior to reproduce the language. This between-group trend continued to T2–T3, where the superior experimental group gains recorded in the earlier stage were maintained, including on T1–T3 variables. DeKeyser (2007) proposed that measurable progress in fluency is made in SA programs of longer duration. It is reasonable to conclude that instruction in combination with the 10-month stay contributed to these positive trends.

SA immersion-only studies often report speech rate gains (e.g., Heunsch & Tracy-Ventura, 2017; Mora & Valls-Ferrer, 2012), including within L2 Mandarin (Di Silvio, Diao, & Donovan, 2016; Du, 2013; Kim et al., 2015). This study confirms but expands these findings to highlight that L2 contact strongly correlated with speech rate advances, showing that learners who are meaningfully active in the L2 and create opportunities to speak Chinese seem to make the greatest gains in fluency. This finding relates to both the control and experimental groups, regardless of any additional instructional benefits. Du (2013) concluded that the amount of time learners spent using Chinese was the most important variable in determining fluency development. These findings for L2 Mandarin are consistent with SA fluency measures of speech rate in research on other languages. Including an examination of time spent using the L2 can shed light on sources of fluency variabilities, but this is yet to be more broadly applied in fluency studies.

**Duration**

Response length, as measured by duration, follows similar patterns to those described so far. Language contact did not correlate with duration either. The T1–T2 comparison showed that the control group made some improvement, but this was not significant, potentially suggesting that such features as breakdown and/or repair strategies were still prevalent, lengthening their overall response times. Without the benefit of instruction on formulaic sequences, the control group perhaps relied on assembling original linguistic components, which may have slowed down the formulation and articulation stages of speech production (Skehan, Foster, & Shum, 2016), since in a L2, the "mental lexicon is smaller, less organised, likely slower in access and contains a narrower repertoire of formulaic language" (Skehan, Foster, & Shum, 2016, p. 98).

Given the formulaic nature of the pre-SA input and that the study's target responses were typically limited to short sequences, analyzing duration of response might at first seem a meaningless exercise. However, we were intrigued by the finding that the experimental group in fact took longer to produce responses at T3 than at T2, seemingly bucking all the positive trends observed in all other variables and time points. Closer examination of the group's actual responses, as reported by Wang and Halenko (2022), suggests that this finding may relate to the longer and more elaborate responses produced at T3 compared with T2, rather than indicating a reversal of fortune where the change could be attributable to an increase in breakdown strategies, for example. It is plausible that the EG utilized the pre-learned knowledge efficiently so that remaining processing capacity was put to more creative use. For example, in the experimental group member's response to Scenario 18 quoted earlier, "老板我是学生, <u>便宜点儿嘛</u>" ("Boss I am student, <u>cheap a bit AUXILIARY</u>"), only the underlined part was the target phrase taught. The other part was the student's own addition, which was creative and appropriate, making the whole response twice as long as the original target phrase. We believe this comparison shows the value of examining data in conjunction with other metrics in order to avoid potential data misinterpretation. Nevertheless, it is noteworthy that the experimental group, despite its move backwards during SA, still outperformed the control group post-SA (at T3) in terms of duration of speech.

A final point to highlight from the entire data set is evidence of individual learner differences, which is by no means unique to this study. Although we were generally able to control for proficiency level and participant profile (e.g., age) between groups, and no significant between-group differences were identified in terms of language contact, we cannot mitigate against other influential personal variables (personality, motivation). It was interesting, however, to observe that the EG displayed fewer within-group individual differences across most fluency variables (as seen in the standard deviations), suggesting that the instruction unified the group's performance, resulting in fluency trajectories that were less diverse than those of the control group.

## Limitations and Future Directions

This study is limited by a small sample size, resulting in potential underpowering. A power analysis using G*Power suggests that a sample size of 24 is needed for adequate power (.80) at an alpha level of .05. Our sample size of 18 is below this threshold, which may limit the ability to detect smaller effects. Additionally, the sample's homogeneity (e.g., students from one university) limits generalizability. The reliance on self-reported L2 contact measures may also introduce bias.

Another limitation is the relatively restricted task type used in this study. The findings may not easily extend to freer oral production tasks, which can offer a broader perspective on fluency. Although we justified not using breakdown and repair measures in this study, these could be valuable in future research with different task types, particularly in examining L2 Mandarin.

Future research could extend the scope to include various task types and other fluency measures, such as breakdown and repair metrics. This study provides a foundation for examining aspects of L2 oral Mandarin, and future research should explore whether these findings can be replicated and extended with different methodologies and broader fluency measures.

## Conclusion

L2 fluency remains "a complex research construct in SLA [second language acquisition], an aspect of performance difficult to define and measure consistently across different tasks and conditions, and a characteristic of language use that many L2 learners may find difficult to develop in and out of the classroom" (Wright & Tavakoli, 2016, p. 73). Much remains to be done to uncover the idiosyncrasies of the SA experience. This longitudinal study found that both the pre-SA intervention and the SA context individually positively impacted fluency development, but a combination of the two proved to be the most successful approach, yielding the greatest impact. The experimental group, which received the intervention, experienced greater gains than the control group both post-instruction and post-SA 1 year later. However, the gap between the experimental group and the control group in terms of planning time and speech rate narrowed during SA, implying the facilitative effect of SA on fluency. The experimental group made progress during this period as well, but not significantly, which might suggest a ceiling effect of the pre-SA instruction for the experimental group during the year abroad. Language contact with the L2 was also found to have a considerable impact on speech rate.

While our study focused on the pragmatic use of language, the findings are applicable to nonpragmatic fluency in several ways. First, improvements in planning time can enhance the ability to formulate speech more quickly and efficiently in any context. Second, an increased speech rate can lead to smoother, more fluent speech in both pragmatic and nonpragmatic settings. Finally, better duration control can help speakers to maintain their speech flow, whether they are engaging in social interactions or delivering a monologue.

Future research may extend the small-scale efforts of this study and its exclusive focus on quantifying aspects of fluency without the support of listener judgments of perceived fluency, for example. Perception studies in L2s other than Mandarin suggest that faster speech rate and mean length of run are clear predictors of higher scores on perceived fluency measures and proficiency levels, whereas slow or disfluent speech has negative effects on the listener (Préfontaine & Kormos, 2016). Perceptions and judgments of speaker fluency can potentially impact the course of interaction, which has wide implications and is in much need of further attention.

Furthermore, given that L2 fluency operates under the influence of a range of cognitive, motivational, social, sociolinguistic, pragmatic, and psycholinguistic variables (Segalowitz, 2016), it is important that research reflects these multidimensional areas going forward. Aspects of motivation (willingness to communicate, beliefs about communication, language, and identity) in particular need further exploration, as learners' cognitive and social experiences of using the L2 can heighten or reduce motivation to communicate and engage, which is the ultimate goal of the SA experience. Instruments measuring language contact, as employed in this study, could effectively incorporate motivational measures to shed light on the impact of these affective dimensions.

This research on SA oral fluency in L2 Mandarin Chinese, a lesser-taught language, significantly enhances the generalizability of findings beyond more commonly studied languages and participant groups. Continued efforts in this direction are crucial for unravelling the complexities of SA outcomes, ultimately leading to more effective and inclusive language learning strategies worldwide.

Final revised version accepted 17 September 2024

## Note

1 Although gender was not considered as a variable in the study design, we performed a nonparametric Mann–Whitney *U* test, which revealed no significant differences between genders ($p = .191$). This test was conducted following reviewer comments, which suggested evaluating potential gender differences in language learning outcomes. However, the analysis indicated that gender did not play a significant role in this context.

## References

Arndt, H. L., Granfeldt, J., & Gullberg, M. (2023). The Lang-Track-App: Open source tools for implementing the experience sampling method in second language acquisition research. *Language Learning*, *73*(3), 869–903. https://doi.org/10.1111/lang.12555

Bardovi-Harlig, K. (2009). Conventional expressions as a pragmalinguistic resource: Recognition and production of conventional expressions in L2 pragmatics. *Language Learning*, *59*(4), 755-795. https://doi.org/10.1111/j.1467-9922.2009.00525.x

Bardovi-Harlig, K., & Bastos, M.-T. (2011). Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics*, *8*, 347–384. https://doi.org/10.1515/iprg.2011.017

Bygate, M. (2001). Effects of task repetition on the structure and CG of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Pearson Longman. https://doi.org/10.4324/9781315838267

Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, *31*(4), 533-557. https://doi.org/10.1017/S0272263109990015

de Jong, N. H., Steinel, M., Florijn, A., Schoonen, R., & Hulstijn, J. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, *34*(1), 5–34. https://doi.org/10.1017/S0272263111000489

DeKeyser, R. (2007). *Practice in a second language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667275

Derwing, T. (2017). L2 fluency development. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 246–259). Routledge. https://doi.org/10.4324/9781315676968

Di Silvio, F., Diao, W., & Donovan, A. (2016). The development of L2 fluency during study abroad: A cross-language study. *The Modern Language Journal*, *100*(3), 610–624. https://doi.org/10.1111/modl.12343

Du, H. (2013). The development of Chinese fluency during study abroad in China. *The Modern Language Journal*, *97*(1), 131–143. https://doi.org/10.1111/j.1540-4781.2013.01434.x

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

François, J., & Albakry, M. (2021). Effect of formulaic sequences on fluency of English learners in standardized speaking tests. *Language Learning & Technology*, *25*(2), 26–41. http://hdl.handle.net/10125/73429

Freed, B. (Ed.). (1995). *Second language acquisition in a study abroad context*. John Benjamins. https://doi.org/10.1075/sibil.9

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. https://doi.org/10.1017/S0272263104262064
García-Amaya, L. (2018). Utterance fluency in the study-abroad context. In C. Sanz & A. Front-Morales (Eds.), *The Routledge handbook of study abroad research and practice* (pp. 166–180). Routledge. https://doi.org/10.4324/9781315639970
Guiaké, M., Chen, D., & Zhang, T. (2021). Meteoric rise of China as a host country for studies: exploring the choice of international students' perspectives. *Journal of Comparative & International Higher Education*, *13*(5), 202-217. https://doi.org/10.32674/jcihe.v13i5.2786
Huensch, A., & Tracy-Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *The Modern Language Journal*, *101*(2), 275-293. https://doi.org/10.1111/modl.12395
Kecskes, I. (2016). Situation-bound utterances in Chinese. *East Asian Pragmatics*, *1*(1), 107–126. https://doi.org/10.1558/eap.v1i1.29098
Kim, J., Dewey, D., Baker-Smemoe, W., Ring, S., Westover, A., & Eggett, D. (2015). L2 development during study abroad in China. *System*, *55*, 123–133. https://doi.org/10.1016/j.system.2015.10.005
Kinginger, C. (2009). *Language learning and study abroad: A critical reading of research*. Palgrave Macmillan. https://doi.org/10.1057/9780230240766
Kinginger, C. (2013). *Social and cultural aspects of language learning in study abroad*. John Benjamins. https://doi.org/10.1075/lllt.37
Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417. https://doi.org/10.1111/j.1467-1770.1990.tb00669.x
Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press. https://doi.org/10.7551/mitpress/6393.001.0001
Li, S. (2014). The effects of different levels of linguistic proficiency on the development of L2 Chinese request production during study abroad. *System*, *45*, 103–116. https://doi.org/10.1016/j.system.2014.05.001
Li, S., Ellis, R., & Kim, J. (2018). The influence of pre-task grammar instruction on L2 learning: An experimental study. *Studies in English Education*, *23*(4), 831–857. https://doi.org/10.1017/S0267190515000069
Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bahtia (Eds.), *Handbook of second language acquisition* (pp. 413–468). New York: Academic Press.
Matsumura, S. (2003). Modelling the relationships among interlanguage pragmatic development, L2 proficiency, and exposure to L2. *Applied linguistics*, *24*(4), 465-491.
Mitchell, R., Tracy-Ventura, N., & McManus, K. (Eds.). (2015). *Social interaction, identity and language learning during residency abroad*. European Second Language Association.

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. https://doi.org/10.1002/tesq.34

Nergis, A. (2021). Can explicit instruction of formulaic sequences enhance L2 oral fluency? *Lingua*, 255, Article 103072. https://doi.org/10.1016/j.lingua.2021.103072

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191–225). Longman. https://doi.org/10.4324/9781315836027

Préfontaine, Y., & Kormos, J. (2016). A qualitative analysis of perceptions of fluency in second language French. *International Review of Applied Linguistics in Language Teaching*, *54*(2), 151–169. https://doi.org/10.1515/iral-2016-9995

Sanz, C. (2014). Contributions of study abroad research to our understanding of SLA processes and outcome. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 1–16). John Benjamins. https://doi.org/10.1075/aals.13

Sanz, C., & Morales-Front, A. (2018). *Routledge handbook of study abroad research and practice*. Routledge. https://doi.org/10.4324/9781315639970

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Segalowitz, N. (2010). Cognitive bases of second language fluenc*y*. Routledge. https://doi.org/10.4324/9780203851357

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. https://doi.org/10.1515/iral-2016-9991

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. https://doi.org/10.1017/S0272263104262027

Skehan, P. (2014). *Processing perspectives on task performance*. John Benjamins. https://doi.org/10.1017/S027226311400076X

Skehan, P., Foster, P., & Shum, S. (2016). Ladders and snakes in second language fluency. *International Review of Applied Linguistics*, *54*(2), 97–111. https://doi.org/10.1515/iral-2016-9992

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G.Widdowson* (pp. 125–144). Oxford University Press. https://doi.org/10.1017/S0272263197223053

Taguchi, N. (2007). Task difficulty in oral speech act production. *Applied Linguistics*, *28*(1), 113–135. https://doi.org/10.1093/applin/aml051

Taguchi, N., Li, S., & Xiao, F (2013). Production of formulaic expressions in L2 Chinese: A developmental investigation in a study abroad context. *Chinese as a*

*Second Language Research Journal*, *2*, 23–58. https://doi.org/10.1515/caslar-2013-0021

Taguchi, N., Li, S., & Xiao, F. (2016). Effects of intercultural competence and social contact on speech act production in a Chinese study abroad context. *The Modern Language Journal*, *100*(4), 775–796. https://doi.org/10.1111/modl.12349

Tavakoli, P., Campbell, C., & McCormack, J. (2016). Development of speech fluency over a short period of time: Effects of pedagogic intervention. *TESOL Quarterly*, *50*(2), 447–471. https://doi.org/10.1002/tesq.244

Tavakoli, P., & Wright, C. (2020). *Second language speech fluency: From research to practice*. Cambridge University Press. https://doi.org/10.1017/9781108589109

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins. https://doi.org/10.1075/lllt.11

Wang, J. (2017). Bringing interculturality into the Chinese-as-a-foreign-language classroom. In T. Jin & F. Dervin (Eds.), *Interculturality in Chinese language education* (pp. 23–44). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-58322-2_2

Wang, J. (2018). British university students studying abroad in China: Issues of interculturality. In F. Dervin, X. Du, & A. Härkönen (Eds.), *International students in China* (pp. 77–94). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-78120-4_4

Wang, J., & Halenko, N. (2019). Longitudinal benefits of pre-departure pragmatics instruction for study abroad: Chinese as a second/foreign language. *East Asian Pragmatics*, *4*(1), 87–111. https://doi.org/10.1558/eap.38216

Wang, J., & Halenko, N. (2022). Developing the use of formulaic language for study abroad: A targeted instructional intervention. *The Language Learning Journal*, *50*(4), 409–426. https://doi.org/10.1080/09571736.2022.2088446

Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *The Canadian Modern Language Review*, *63*(1), 13–33. https://doi.org/10.3138/cmlr.63.1.13

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wright, C. (2021). Effects of task type on L2 Mandarin fluency development. *Journal of Second Language Studies*, *3*(2), 157–179. https://doi.org/10.1075/jsls.00010.wri

Wright, C., & Tavakoli, P. (2016). New directions and developments in defining, analyzing and measuring L2 speech fluency. *International Review of Applied Linguistics*, *54*(2), 73–77. http://dx.doi.org/10.1515/iral-2016-9990

Yan, X. (2020). Unpacking the relationship between formulaic sequences and speech fluency on elicited imitation tasks: Proficiency level, sentence length, and fluency dimensions. *TESOL Quarterly*, *54*(2), 460–487. https://doi.org/10.1002/tesq.556

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*(2), 1–27. https://doi.org/10.1093/applin/24.1.1

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Accessible Summary**
**Appendix S1.** Twenty-Six Scenarios and Target Responses.
**Appendix S2.** Normality Test Results.
**Appendix S3.** Language Contact Profile.