


RESEARCH ARTICLE

A longitudinal investigation of the effects of language instruction versus immersion on cognitive functions in young adult Chinese speakers learning English

Lihua Xia^{1,2,*} , Thomas H. Bak², Mariana Vega-Mendoza^{2,3} and Antonella Sorace²

¹Huazhong University of Science and Technology, Wuhan, China; ²The University of Edinburgh, Edinburgh, Scotland; ³Luleå University of Technology, Luleå, Sweden

*Corresponding author. E-mail: lihuaxia@hust.edu.cn

(Received 17 December 2020; Revised 11 March 2022; Accepted 14 March 2022)

Abstract

The current study examined cognitive effects of two pathways of second language (L2) acquisition longitudinally in Chinese speakers learning English in an L2-dominant environment. Thirty-nine participants who attended an intensive 10-week English course (L2-instruction group) were compared to 38 participants who attended regular university courses taught in English (L2-immersion group). Four repeated assessments were conducted over 10 weeks: precourse (baseline) and postcourse assessments, and two interim assessments every 3 weeks. Both groups matched on background variables (e.g., intelligence) and showed comparable cognitive performance in all measures at the baseline. The longitudinal results showed a similar improvement in both groups for most cognitive measures, such as visual and auditory inhibition. The only significant group difference was observed in the auditory inhibition test, where the L2-instruction group outperformed the L2-immersion group. Taken together, our results suggest a specific effect of language experience and an overall effect of linguistic context on cognitive functions.

Introduction

The term “immersion education” was adopted in Canada during the 1960s to describe the programs where English-speaking children were taught through the medium of French (Cummins, 1998; Genesee, 2006). Nowadays, the term is used to describe the approach to the second language (L2) instruction, in which the usual curricular activities are conducted in the L2. There are different types of language immersion programs, such as Content and Language Integrated Learning (CLIL) popular in most European countries (De Graaff et al., 2007) or Study Abroad Programs (Dewey, 2004).

Studies that examined the effects of immersion education on linguistic performance have reached a relatively uniform agreement that immersion education conducted in an L2-dominant linguistic environment (e.g., study abroad) significantly improves L2 speaking proficiency (Hernández, 2010; Linck et al., 2009; Llanes & Muñoz, 2009;

Segalowitz *et al.*, 2004; Serrano *et al.*, 2011; Tanaka & Ellis, 2003). In contrast, findings for other linguistic domains have been inconsistent (for a review see Pliatsikas & Chondrogianni, 2015). There is increasing evidence that L2-learners who received traditional classroom-based L2 instruction outperformed those receiving L2-immersion instruction in linguistic abilities, such as grammatical and lexical abilities (Collentine, 2008; Parafita Couto *et al.*, 2015), pronunciation of selected features of phonology (Segalowitz *et al.*, 2004), and verbal learning (Dahl & Vulchanova, 2014). However, L2-immersion instruction was reported to positively correlate with performance on the lexical categorization task (Malt & Sloma, 2003), while years of classroom-based L2 training was negatively associated with lexical performance (Zinszer *et al.*, 2014). This inconsistency may be attributed to a range of variables such as self-confidence and attitudes (Serrano, 2011; White & Turner, 2005), L2 proficiency (Hernández, 2010; Tanaka & Ellis, 2003), length of L2 immersion (Malt & Sloma, 2003), age of earliest L2 exposure, code-switching patterns, classroom instruction (Zinszer *et al.*, 2014), and focus of L2 instruction (Collentine, 2008). In addition, it has been shown that the manner in which learners acquire their L2 (e.g., intensive vs. regular learning style) is highly correlated with L2 linguistic performance; intensive L2 input and exposure predict better L2 competence (Collins *et al.*, 1999; Freed *et al.*, 2004; Serrano *et al.*, 2011). However, little is known whether these two pathways of language acquisition (i.e., L2 instruction vs. immersion) might have a different impact on the cognitive effects associated with language learning.

As for the cognitive consequences of language immersion, most studies have examined these cognitive effects in children who were enrolled in immersion education, which provides a nativelike linguistic environment. Prior research has suggested a positive effect of L2 immersion on cognitive control. For example, Nicolay and Poncet (2013) compared the cognitive performance of children enrolled in English immersion classes for 3 years and those enrolled in monolingual French-speaking classes. The results demonstrated that L2-immersion positively influenced attentional/executive control, such as alerting, auditory selective attention, divided attention, and mental flexibility, but not interference inhibition. Their follow-up study (i.e., Nicolay & Poncet, 2015) used a longitudinal design, and the results were consistent with their previous findings. Similar studies on benefits associated with language immersion have been reported with respect to general cognitive abilities, such as intelligence (Woumans *et al.*, 2016), verbal working memory, and word learning (Kaushanskaya *et al.*, 2014). These cognitive benefits were correlated to the level of L2 proficiency and length of immersion (Bialystok & Barac, 2012).

However, other studies reported no cognitive effects of L2 immersion in children, suggesting that these effects might be confounded by language proficiency (Poarch & van Hell, 2012) or other types of experiences (Carlson & Meltzoff, 2008), such as musical experience (Janus *et al.*, 2016), video-gaming experience (Vulchanova *et al.*, 2015), physical activities (e.g., martial arts, training in aerobics, and yoga), and games (Diamond 2012; Diamond & Lee, 2011). Simonis *et al.* (2020) found that the unmatched background variables between groups (i.e., age, socioeconomic status [SES], and intelligence) could lead to a null result when assessing the cognitive differences through an across-group comparison. In addition, these confounding variables are complex and interact with each other. For example, a study with Chinese-English bilinguals showed that L2 proficiency along with SES and general intelligence were strong predictors of performance on conflict monitoring (Xie & Pisano, 2018). Given that the bilingual experience itself is fluid, complex, and dynamic (DeLuca *et al.*, 2019; Luk & Bialystok, 2013), there are some experience-based factors (e.g., linguistic environment of L2 acquisition, L2 exposure, and L2 proficiency) that might

lead to different linguistic and cognitive consequences. Therefore, a longitudinal design has been strongly recommended to assess the cognitive effects associated with language immersion (Nicolay & Poncelet, 2013, 2015; Simonis et al., 2020; Woumans et al., 2016), which allows for correlations between experience-based factors and language immersion-induced cognitive changes.

Study abroad, a specific immersion education program, provides a naturalistic linguistic environment, which allows L2-learners to have more social interaction in L2 and dramatically enhances L2 input in terms of quantity and quality (DeKeyser, 2007). The Adaptive Control Hypothesis (ACH; Green & Abutalebi, 2013) has been proposed to describe how cognitive control processes adapt to the linguistic contexts, in which individuals are engaged with different modes of language interaction (i.e., single language, dual language, and dense code-switching). For example, in a single language context, the cognitive demands on interference control (i.e., conflict monitoring and interference suppression) are significantly increased in L2 learners for selecting the target language while inhibiting the nontarget language. Specifically, in a single and L2-dominant linguistic environment, L2 learners are hypothesized to rely more on interference control to adjust to the linguistic environment. More empirical evidence has shown that the linguistic environment has impact on individuals in both monolinguals and bilinguals. For instance, monolinguals who lived in a multilingual environment compared to those living in a monolingual environment showed enhanced brain's receptivity to learning a new language (Bice & Kroll, 2019). Another study found changes in the structure of a dynamic system (i.e., the cerebellum, forceps minor, caudate, and hippocampus) that showed an overall adaptation within the language control networks in adult sequential bilinguals, who did not receive any linguistic learning or training while living in the L2-dominant environment (DeLuca et al., 2019). These observations suggest that the linguistic context (i.e., an L2-dominant linguistic environment) plays an important modulating role in individuals' language learning and cognitive processes in relation to language usage and control.

Given the limited research on the cognitive consequences of different pathways of L2 acquisition (i.e., L2-instruction vs. L2-immersion) in young adults within an L2-dominant linguistic environment, the current study investigates these cognitive effects while considering the theoretical and methodological aspects of previous research. Specifically, we aim to explore (1) whether different pathways of language acquisition affect cognitive control differently by comparing performance for cognitive tasks in two groups of Chinese-English bilinguals who continued their L2 learning through instruction and immersion education, respectively; and (2) to what extent the L2-dominant linguistic context modulates these cognitive effects through a longitudinal design to track the potential cognitive changes induced by different language experiences and a naturalistic linguistic environment. The instruction and immersion groups were comparable on background measures (e.g., age, SES, and intelligence). To avoid the issue of task impurity (Miyake & Friedman, 2012) and limitations associated with the use of a single indicator to measure a single aspect of cognitive functions (von Bastian et al., 2016), we adopted a testing battery comprised of a series of theoretically motivated and commonly used experimental tasks that engage specific aspects of cognitive control (e.g., inhibitory control) in both the visual and auditory domains. The selection of the nonlinguistic cognitive tasks was based on the fact that the nature of language learning entails multiple domains, such as visual (e.g., reading and writing), auditory (e.g., listening and speaking), and working memory (e.g., language processing).

A novel aspect of the study is the use of a longitudinal design with four repeated assessments. This design has the following advantages: (1) the traditional cross-sectional

comparative (e.g., immersion vs. nonimmersion) designs may not fully capture the effects of experience-based factors (e.g., language exposure and linguistic environment) on cognitive changes induced by language immersion (Deluca *et al.*, 2019), whereas a longitudinal design can examine how specific experience-based factors are reflected in executive functions, and crucially how they change over time; and (2) previous longitudinal design normally uses two testing points (e.g., pretest vs. posttest), the design of four testing points is relatively stable and reliable to trace within-subject changes, which can provide a better understanding of changes in cognitive performance over time due to the two pathways of language acquisition as well as L2-dominant linguistic environment. The main potential concern of this design, however, would be a practice effect (see “Discussion” section).

Method

Participants

Two groups of Chinese students were recruited from the University of Edinburgh. None of them had lived in any country other than China before and they participated in this study just after they arrived at Edinburgh; thus, the length of naturalistic exposure to English during the testing session was the same. One group of participants ($n = 39$) were prospective master students who attended a 10-week Pre-session English (PSE) course (i.e., L2-instruction group) in June, starting prior to the university master courses in autumn. The L2-instruction group received an intensive language training course. The second group of participants ($n = 38$) did not attend the language course (i.e., L2-immersion group) but started directly to attend regular subject-based academic courses taught in English in September. The L2-immersion group received a specific “immersion education.” There were three and seven participants from the College of Science and Engineering in the L2-instruction and L2-immersion group, respectively, and the others were from the College of Arts, Humanities, and Social Science. All participants were native speakers of Mandarin Chinese and English was their main second language (L2). In the L2-instruction group, 11 participants reported having learned an L3: French (2), Japanese (8), and Korean (1), and 4 participants reported having learned an L4: Cantonese (1), French (1), German (1), and Korean (1). In the L2-immersion group, 20 participants reported having learned an L3: French (7), German (1), Japanese (8), Korean (3), and Spanish (1), and one participant reported to have learned Spanish as L4. The self-reported proficiencies in L3s and L4s were very low: average scores are 0.96 and 0.46, respectively (scale: 0–5). Two participants in each group did not attend the last session of testing, but their data from the prior three testing sessions were included in the analyses. Participants’ demographic information is presented in [Table 1](#).

The 10-week PSE is a course for prospective students preparing to enter a post-graduate degree program, which takes place in summer prior to the university’s opening date. Students who attend the PSE usually have not reached the language requirement for their master programs. The PSE program consists of two parts: English for Academic Purposes (EAP; 6 weeks) and English for Specific Academic Purposes (ESAP; 4 weeks). The EAP is divided into two 3-week courses (EAP1, weeks 0–3; EAP2, weeks 3–6), which focus on writing (paraphrasing, summarizing, and synthesizing), listening (lecture listening and note-taking), speaking (discussion and scenarios), reading (reading academic texts), and academic language study. The ESAP particularly focuses on the development of subject-based academic writing and presentation skills of specific subjects. The course consists of four 50-minute morning classes every

Table 1. Demographic information of participants. SDs are given in parentheses

	L2-instruction group	L2-immersion group
<i>N / male</i>	39 / 2	38 / 2
<i>Raven's APM scores</i> ¹	17.87 (3.25)	19.26 (2.99)
<i>Age in years</i>	22.41 (1.07)	22.39 (0.82)
<i>SES indexed by parents' education</i> ²	7.19 (1.16)	6.76 (1.8)
<i>Age of L2 acquisition in years</i>	8.38 (3.28)	7.47 (2.37)
<i>Musical experience %</i>	64.10%	55.26%
<i>Video-gaming Experience %</i>	56.41%	60.53%
<i>Frequency of language use</i>		
<i>Childhood</i>	2.77 (1.42)	2.84 (1.52)
<i>Adulthood</i>	3.54 (1.10)	4.18 (0.93)*
<i>Last two years</i>	3.95 (0.92)	4.34 (0.78)*
<i>Average hours of practicing L2</i> ³	7.76 (2.60)	7.06 (2.81)
<i>Average hours of classes</i>	19.6 (3.34)	12.78 (5.54)*
<i>IELTS overall scores</i> ⁴	6.45 (0.25)	7.04 (0.34)*
<i>Speaking</i>	5.92 (0.35)	6.26 (0.36)*
<i>Listening</i>	6.78 (0.58)	7.51 (0.75)*
<i>Reading</i>	6.92 (0.48)	7.76 (0.67)*
<i>Writing</i>	5.96 (0.27)	6.28 (0.41)*
<i>Self-reported L2 proficiency at W0</i> ⁵	11.1 (2.1)	12.47 (1.84)*
<i>Speaking</i>	2.54 (0.72)	2.84 (0.68)
<i>Understanding</i>	2.97 (0.71)	3.26 (0.69)
<i>Reading</i>	3.13 (0.57)	3.68 (0.57)*
<i>Writing</i>	2.46 (0.68)	2.82 (0.56)*
<i>Self-reported L2 proficiency at W10</i>	12.51 (2.14)	13.11 (2.44)
<i>Speaking</i>	2.84 (0.6)	2.97 (0.77)
<i>Understanding</i>	3.16 (0.65)	3.42 (0.73)
<i>Reading</i>	3.38 (0.68)	3.64 (0.72)
<i>Writing</i>	3.14 (0.75)	3.08 (0.6)

¹APM scores were the number of corrected items (the total number was 36).

²This a composite score based on parental education level given by the summed scores of both parents. The scale ranged from 1: *primary school*, 2: *middle school*, 3: *high school*, 4: *bachelor's or equivalent to*, to 5: *postgraduate*. Frequency of language use: *never*, 1; *yearly*, 2; *monthly*, 3; *weekly*, 4; *daily*, 5

³The average hours of practicing L2 per week over the 10 weeks, activities include watching TV, classes, reading books, conversation in L2, listening to radio, reading, and internet.

⁴IELTS is an international standardized test of English language proficiency for nonnative English language speakers. The score ranged from 0 to 9 (high score means better performance), in 0.5 band increments.

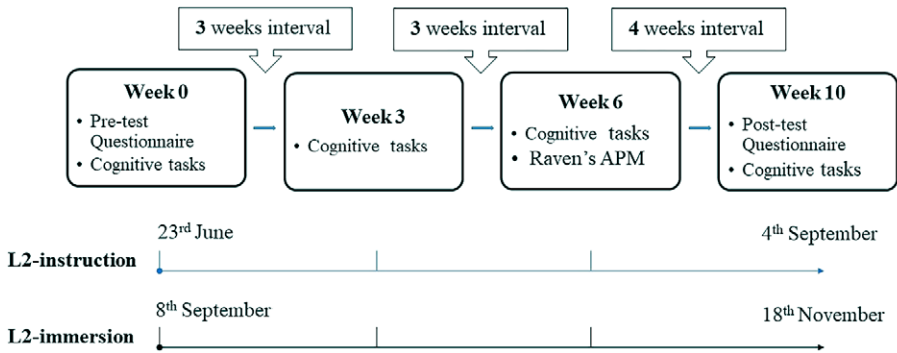
⁵The scale of self-reported L2 proficiency ranged from 0–5, marked by “*none*” to “*fluent*.”

*Significant differences between L2-instruction group and L2-immersion group.

working day. Based on the design of the EAP language program, we conducted longitudinal testing across four sessions: at the beginning of language course (W0, baseline), end of EAP1 (W3), end of EAP2 (W6), and end of ESAP (W10). Both groups received the same procedure of testing with four repeated sessions, but the L2-instruction group started earlier than the L2-immersion group (see Figure 1).¹

¹The Raven's APM was administered at W6 instead of W0 for the following reasons: (1) this task was used as a measure of nonverbal general intelligence, which is assumed to be stable in adulthood in a short time (i.e., 6 weeks; Schaie, 1996). There is no strong reason to believe that the two groups of participants would show different improvements in terms of IQ in 6 weeks; and (2) the time of the testing session and the rate of return for this longitudinal design: It was the participants' first time in taking part in such kind of psychological research; thus, it took much longer than the expected time (i.e., 1.5h).

a. Timeline of four repeated sessions



b. Cognitive tasks procedure

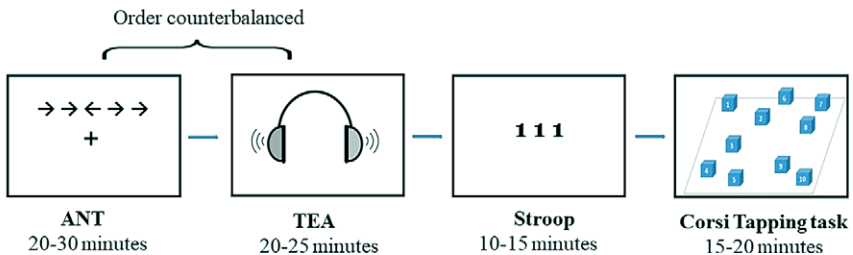


Figure 1. Timeline of the four repeated sessions (a) and the procedure of cognitive tasks (b).

Background measures

The Raven’s Advanced Progressive Matrices

The Advanced Progressive Matrices (APM) (Raven & Foulds, 1962) was used as a control of nonverbal general intelligence. Consistent with a previous study (Xia *et al.*, 2022), we adopted Set I (i.e., Item 5 or Item 7) as practice and Set II as the experimental test. The design of the matrices ensures that the demand of the level gradually increases with the items. Participants were instructed to complete the matrices item by item in 10 minutes and were told that if they were having difficulty with a specific item, they could guess the answer and continue to the next one. They started with Item 1 and had to answer as many items as they could. The results were scored as the number of correct items for each participant.

Questionnaires

Two English version of the questionnaires were employed during the first and the last sessions, respectively. The pretest questionnaire was to collect general demographic information (e.g., age and gender), IELTS (i.e., International English Language Testing System) scores, and self-reported language proficiency before the testing. Using 6-point scales (i.e., 0–5, marked from “none” to “fluent”), participants rated their speaking, understanding, reading, and writing skills in every language they had learned. Information on other confounding variables that might affect executive performance was

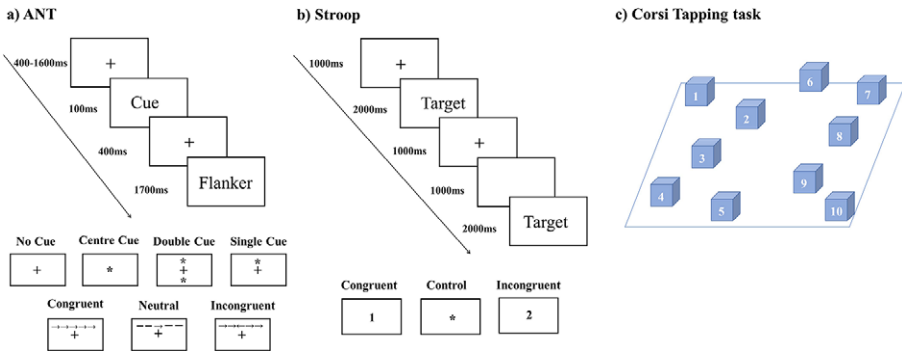


Figure 2. Schematic representation of each task of executive functions.

also collected, including the age of L2 acquisition (AoA), parents’ education (as an index of SES), musical experience, and video-gaming experience. The posttest questionnaire was to collect the information on their self-reported English proficiency after 10 weeks, language usage, and the time they spent practicing/learning the L2.

Experimental tasks

Four nonlinguistic cognitive tasks were employed to measure different aspects of executive functions. In the computerized tasks, all stimuli were presented with E-Prime (version 2.0) on a 17-inch computer screen. A schematic representation of each task is depicted in Figure 2.

Attention Network Task

This task is a well-established assessment of attentional capacities (i.e., alerting, orienting, and inhibition) (Fan et al., 2002), which has been used to investigate the effects of bilingualism on attentional abilities (e.g., Costa et al., 2008). We used this task to investigate the impact of L2 instruction and linguistic context on these attentional capacities. In this task, participants were instructed to respond to the central arrow of the horizontal five arrows presented in the middle of the screen either below or above a fixation cross. There were three types of trials: congruent, neutral, incongruent, and four cueing conditions: single, double, center, and no cue. Three attentional indices were obtained calculating the difference in RTs/accuracy rate between the following trials: Attention Network Task (ANT) conflict (congruent vs. incongruent); ANT alerting (double-cue vs. no-cue); ANT orienting (center-cue vs. single-cue). Participants started with a practice block consisting of 24 trials and followed by three experiment blocks of 96 trials each. Feedback on performance was only provided in the practice block.

Number Stroop task

To avoid any linguistic influence, a numerical version of the Stroop task was used (adapted from Hernández et al., 2010). This task is well documented to measure inhibitory control (i.e., Stroop effect) (Stroop, 1935) and is the primary focus of the current study. In this task, participants were asked to count digits or symbols presented

on the center of the screen by pressing the keys 1, 2, or 3 on the keyboard while ignoring the numerical value of the digits. There were three experimental conditions: congruent, incongruent, and neutral condition. Inhibitory control was assessed by the difference in RTs/accuracy rate between incongruent and congruent trials. Participants were given a practice block with 18 trials and followed by two experimental blocks of 90 trials each. Feedback on performance was only provided in the practice block.

Test of Everyday Attention

The Test of Everyday Attention (TEA) (Robertson *et al.*, 1994) is a well-established clinical assessment of attention, which has three parallel versions (*i.e.*, A, B, and C) to avoid practice effects. Most studies examining the cognitive effects associated with language learning/bilingualism have focused on the visual attentional domain. We selected the three subtests of Elevator Tasks to explore these effects in the auditory domain. To avoid practice effects, the three parallel versions were used in the order of A-B-C-A in the current study. All tasks were presented through media player with a headset.

- (a) Elevator with Counting (EC: 7 trials): This task assesses sustained attention. Participants were asked to count tones of the same pitch presented at irregular intervals.
- (b) Elevator with Distraction (ED: 10 trials): This task assesses auditory selective attention/inhibition. Participants were asked to count low tones while ignoring interspersed high tones.
- (c) Elevator with Reversal (ER: 10 trials): This task assesses auditory attentional switching (auditory-verbal working memory). Participants were presented with high, middle, and low tones. They had to count the middle tones while the high and low tones indicated the counting direction (upward and downward, respectively).

Corsi Tapping Task

The Corsi Tapping Task (CTT) was adapted from the Wechsler Memory Scale-III (WMS-III, Wechsler, 1997), which was established as a measure of visuospatial memory span (forward condition), working memory (backward condition), and mental rotation (rotated condition). Working memory plays a crucial role in L2 acquisition (Ardila, 2003). Previous studies have shown that the management of two languages results in more efficient executive processing, including working memory (Grundy & Timmer, 2017) and spatial reasoning (Greenberg *et al.*, 2013). This task began with the simple forward condition and ended with the difficult rotated condition. It was presented on a plastic whiteboard (27.5 cm × 21cm) with 10 blue cube-shaped numbered blocks (3 cm × 3 cm; from 1 to 10), but the numbers were only visible to the experimenter. The experimenter tapped on blocks sequences at a rate of approximately 1 second per block, with sequences varying in length from two to nine blocks. Participants reproduced the sequences in their original order (beginning with two two-block sequences) in the forward condition and reversed the order in the backward condition. The tasks stopped when participants made errors on both sequences at a given length (*e.g.*, six-block sequences).

In addition to forward and backward conditions (Bialystok *et al.*, 2008), the rotated condition was used to measure mental rotation (adapted from Keehner & Gathercole, 2007). In this condition, there were two identical whiteboards, with numbers facing the experimenter only and one of the boards was rotated for 180° from the other one. The

experimenter tapped on the blocks in a predetermined sequence, and the participants had to reproduce the sequences with the same blocks and the same order on the rotated board. The sequences varied in length from one to five blocks. The task stopped when the participants make errors on all six sequences at a given length. This task started with four one-block practice trials, and there were no practice trials in the forward and backward conditions. As for scoring, participants got one point for each correctly tapping trial, and results were scored as the percentage of accuracy based on possible number of trials.

Statistical analyses

All analyses were conducted using the *lme4* packages (Bates et al., 2015) in R (Version 3.6.1, R Core Team, 2019) and the *lmerTest* package (Kuznetsova et al., 2017) was used to compute *p*-values and conduct pairwise comparisons to test significant interactions.

In the initial analysis, background variables with a continuous scale (e.g., age and SES) were analyzed using one-way ANOVA, while other experience measures on a nominal scale (e.g., gender) were using the chi-squared test. The factors that indicated group differences were added to models in the main analyses. In the main analyses of the RT-based tasks, the linear mixed-effect models (LMMs) were employed. The main motivation for using LMMs in favor of traditional analysis is that LMMs are robust against unbalanced datasets and based on by-subjects and by-items analyses (Baayen et al., 2008). Moreover, LMMs offer additional benefits to SLA researchers: It gives more weight to participants with more data; it can change over time according to how L2 acquisition develops; and it includes all raw data and has more statistical power (Cunnings, 2012). Before data analyses, following Xia et al. (2022), we excluded the trials for which RTs were outside of 3-SD of each participant mean across all trial types and RTs associated with incorrect responses. Thus, the total trials excluded in each task were the following: 2.87% and 4.32 % trials in the ANT and Stroop, respectively.

In the LMMs, RTs was the dependent variable, Group (i.e., L2-instruction group vs. L2-immersion group), Session (i.e., four repeated sessions: W0, W3, W6, and W10), and Trial Type (e.g., congruent and incongruent trials in the model for the conflict effects) were fixed variables unless specified, and participants and items were random variables (i.e., including random intercepts for each participant and item). Fixed variables were allowed to interact with each other in a single model. The Trial Type was varied according to the effects of interest; thus, there were four models in the ANT (i.e., Overall RTs, Conflict, Alerting, and Orienting) and two models in the Stroop (i.e., Overall RTs and Stroop). Another LMM was run with the fixed variables that indicated significant main/interaction effects to demonstrate the differences between different levels of the fixed variables. Because the accuracy rate was relatively high in the RT-based tasks (i.e., ANT: 98.71%; Stroop: 97.57%), therefore, we only analyzed the average accuracy rate across the four sessions. Only the interactions were reported, which was for reasons of brevity and relevance to the research questions. For the three subtests of the TEA and the CTT, the accuracy rate was obtained based on the number of correct responses, therefore linear regression models (LRs) were used with the accuracy rate as a dependent variable, Group and Session as fixed variables. In the analysis of self-reported English proficiency, LRs were used with the proficiency scores as a dependent variable, Group and Testing Time (i.e., pretest vs. posttest) as fixed variables. The fixed and random effects outputs of models are presented in supporting materials.

We checked the model by removing/adding slopes to select the best-fitted models (Baayen *et al.*, 2008). The results showed that only adding slope for each participant fitted the best model. Because there were four testing points, we added a quadratic term to the model to allow us to test whether the effect of session became progressively stronger or leveled off (Steinberg *et al.*, 2018). We checked to add a quadratic term or not and found that it did not change the results. To make our model simple and easy to be converted, we excluded the quadratic term from the main analyses. Due to the multiple levels of the session (*i.e.*, 4 levels), following Spronken *et al.* (2016), we used the *anova* function to return *F* statistics corresponding to the sequential decomposition of the overall effects of fixed variables, and the *summary* function to return *t* statistics corresponding to the comparisons between different levels of fixed variables as well as their interactions (Bates *et al.*, 2015; Kuznetsova *et al.*, 2017). We additionally calculated pseudo R^2 values using the *r.squaredGLMM* function of the package MuMIn for the LMMs (Barton, 2014), which provides a value for marginal R^2 (variance explained by fixed effects) and a value for conditional R^2 (variance explained by both fixed and random effects). We used the *r.squaredLR* function for the LRs, which provides a value for R^2 and a value for adjusted R^2 .

Results

Initial analyses

Background measures

No group differences were found on age, age of attending primary school, SES, AoA, musical experience, video-gaming experience, and Raven's APM scores (all $ps > .05$), indicating comparable fundamental cognitive abilities in the two groups (see Table 1).

Language-related variables

The two groups did not differ on average hours of practicing L2 over the 10 weeks ($p > .05$). Due to the design of the PSE program, the class time (*i.e.*, formal English-instruction classes in the L2-instruction group vs. classes of academic program conducted in English in the L2-immersion group) was longer in the L2-instruction group than the L2-immersion group ($p < .001$). As to the IELTS, as expected, the L2-immersion group obtained significantly higher scores relative to the L2-instruction group on overall scores as well as subcategories (*i.e.*, speaking, listening, reading, and writing) (all $ps < .05$). In terms of previous frequency of language use, the L2-immersion group showed higher frequency of language use in the adulthood and in the last two years than that of the L2-instruction group (all $ps < .05$), which is consistent with the group difference in language proficiency. In the analysis of self-reported L2 proficiency, the main effect of group was significant: The L2-immersion group showed higher scores than L2-instruction group on the overall proficiency ($Est = 0.98$, 95% CI [0.29, 1.67], $t = 2.82$, $p = .005$). The main effect of Testing Time was significant: significant increase on overall self-reported proficiency at posttest ($Est = 1.02$, 95% CI [0.33, 1.71], $t = 2.94$, $p = .004$). The interaction between Group and Testing Time was not significant ($p = .27$) ($R^2 = .11$, $R^2_{adjusted} = .09$).

Prestage analyses

To investigate the effects of the L2 experience (*i.e.*, immersion vs. instruction), we compared the performance of two groups at the first testing point (*i.e.*, W0) and did not

observe group differences in any of the cognitive measures (all $ps > .05$). To longitudinally explore when the changes occur and examine the degree of change over the four repeated sessions, we compared the performance between testing points (i.e., W0–W3 vs. W3–W6 vs. W6–W10). These analyses were run by changing the reference level in the original models from “W0” to “W3” and “W6,” respectively. Changes were observed between the testing points W0–W3 in most measures (details see “Main Analyses”), and no changes were found between the testing points W3–W6 and W6–W10 in any of the cognitive measures (all $ps > .05$). These findings indicated that these observed changes occurred at the earlier stage of testing session and lasted until the last session.

Main analyses

Attention Network Task

Mean RTs on the respective trial types are given in Table 2.

Overall performance

No main effects or interaction effects were found on the overall RTs (*marginal* $R^2 = .003$, *conditional* $R^2 = .38$) and accuracy performance (all $ps > .05$), indicating that there was no practice effect.

Alerting effect

The Alerting effect was significant, indicating faster responses on double-cue trials than on no-cue trials ($Est = 55.66$, 95% CI [28.66, 82.66], $t = 4.04$, $p < .001$). No other main effects or interactions were significant (all $ps > .05$) (*marginal* $R^2 = .063$, *conditional* $R^2 = .376$).

Orienting effect

The Orienting effect was significant, indicating faster responses on single-cue trials than on center-cue trials ($Est = 30.02$, 95% CI [1.65, 58.39], $t = 2.07$, $p = .038$). No other main effects or interactions were significant (all $ps > .05$) (*marginal* $R^2 = .022$, *conditional* $R^2 = .375$).

Table 2. ANT indices by Sessions. SDs are given in parentheses

	L2-instruction group				L2-immersion group			
	W0	W3	W6	W10	W0	W3	W6	W10
Congruent	456 (53)	447 (54)	450 (47)	447 (59)	453 (56)	445 (51)	446 (52)	448 (57)
Incongruent	540 (61)	514 (61)	510 (53)	505 (62)	531 (60)	513 (64)	506 (60)	501 (62)
Conflict	84 (22)	67 (21)	60 (18)	58 (16)	77 (24)	68 (25)	60 (23)	53 (21)
Double-cue	473 (52)	455 (51)	454 (41)	450 (52)	467 (53)	457 (55)	452 (50)	453 (56)
No-cue	524 (58)	513 (62)	513 (47)	511 (64)	517 (62)	513 (59)	507 (54)	512 (66)
Alerting	51 (21)	57 (23)	59 (17)	60 (26)	49 (24)	55 (24)	55 (19)	59 (30)
Single-cue	448 (49)	436 (51)	438 (48)	435 (58)	447 (56)	438 (58)	434 (56)	433 (60)
Center-cue	483 (56)	465 (53)	464 (51)	460 (57)	479 (62)	465 (55)	463 (56)	466 (64)
Orienting	35 (16)	29 (15)	26 (17)	25 (19)	33 (21)	27 (15)	29 (16)	33 (20)

Note: Conflict: incongruent vs. congruent trials; Alerting: no-cue vs. double-cue trials; Orienting: center-cue vs. single-cue trials.

Conflict effect

The Conflict effect was significant, indicating faster responses on congruent trials than on incongruent trials ($Est = 66.30$, 95% CI [45.47, 87.12], $t = 6.24$, $p < .001$). The interaction between Conflict effect and Session was significant, indicating improvements in inhibition with testing sessions starting from W3 (W3 vs. W0: $Est = -13.35$, 95% CI [-20.05, -6.64], $t = -3.90$, $p < .001$; W6 vs. W0: $Est = -20.88$, 95% CI [-27.59, -14.17], $t = -6.10$, $p < .001$; W10 vs. W0: $Est = -25.05$, 95% CI [-31.85, -18.25], $t = -7.22$, $p < .001$). No other main effects or interactions were significant (all $ps > .05$) ($marginal R^2 = .090$, $conditional R^2 = .4$) (see Figure 3).

Number Stroop Task

Mean RTs on the respective trial types are given in Table 3.

Overall Performance

The overall accuracy did not change but mean RTs decreased over testing sessions, starting from W3 (W3 vs. W0: $Est = -43.74$, 95% CI [-67.63, -19.86], $t = -3.59$, $p < .001$; W6 vs. W0: $Est = -56.95$, 95% CI [-80.84, -33.06], $t = -4.67$, $p < .001$; W10 vs. W0: $Est = -59.58$, 95% CI [-83.88, -35.28], $t = -4.81$, $p < .001$). No other main effects or interactions were significant (all $ps > .05$) ($marginal R^2 = .035$, $conditional R^2 = .426$).

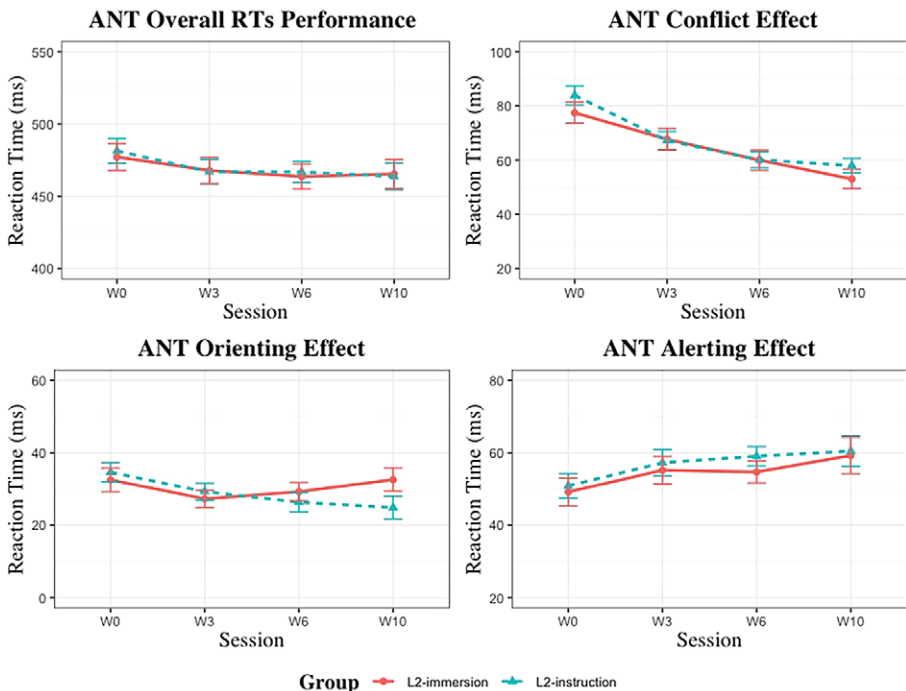


Figure 3. Performance in the ANT over the four repeated assessments by group. Error bars represented $\pm 1SE$.

Table 3. Stroop indices by sessions. SDs are given in parentheses

	L2-instruction group				L2-immersion group			
	W0	W3	W6	W10	W0	W3	W6	W10
Congruent	530 (72)	498 (75)	486 (69)	483 (66)	535 (80)	492 (60)	479 (67)	480 (72)
Incongruent	606 (82)	559 (89)	548 (88)	540 (87)	615 (95)	547 (73)	533 (86)	529 (85)
Stroop	76 (28)	61 (34)	62 (34)	57 (30)	80 (32)	55 (24)	54 (31)	49 (23)

Note: Stroop effect: incongruent vs. congruent trials.

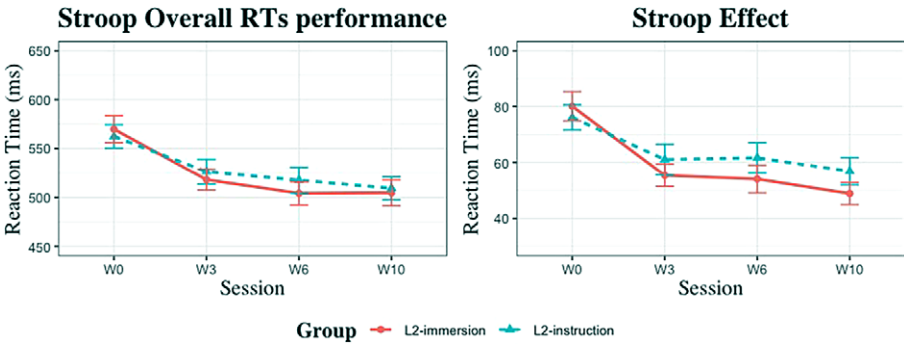


Figure 4. Performance in the Number Stroop task over the four repeated assessments by group. Error bars represent $\pm 1SE$.

Stroop effect

There were main effects of Stroop effect [$F(1, 7.22) = 44.98, p < .001$] and Session [$F(3, 295.01) = 11.00, p < .001$], and a significant interaction between them [$F(3, 294.30) = 10.90, p < .001$]: improvements in inhibitory control with testing sessions starting from W3 (W3 vs. W0: $Est = -20.06, 95\% CI [-29.46, -10.65], t = -4.18, p < .001$; W6 vs. W0: $Est = -20.43, 95\% CI [-29.84, -11.03], t = -4.26, p < .001$; W10 vs. W0: $Est = -25.40, 95\% CI [-34.96, -15.83], t = -5.20, p < .001$). No other main effect or interactions were significant (all $ps > .05$) ($marginal R^2 = .094, conditional R^2 = .446$) (see Figure 4).

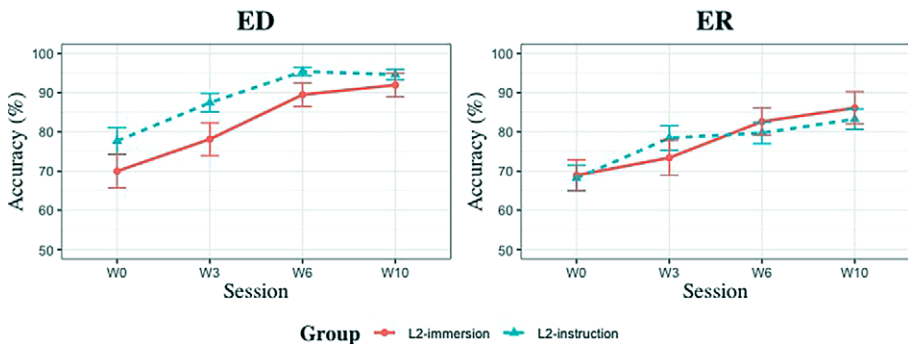
Test of Everyday Attention

Performance on each sub-task of the TEA is given in Table 4.

In the EC, no main effects or interaction effects were found (all $ps > .05$) ($R^2 = .026, R^2_{adjusted} = .003$). The average accuracy rate of the two groups was 98.40%, indicating a ceiling effect on performance. In the ED, the main effect of Group was significant: the L2-instruction group performed better on selective attention/inhibition than the L2-immersion group did ($Est = -6.43, 95\% CI [-11, -1.87], t = -2.77, p = .006$); the main effect of Session was significant: increased performance with testing sessions from W3 (W3 vs. W0: $Est = 6.43, 95\% CI [1.87, 11], t = 2.77, p = .006$; W6 vs. W0: $Est = 18.57, 95\% CI [12.56, 24.58], t = 6.08, p < .001$; W10 vs. W0: $Est = 19.39, 95\% CI [13.3, 25.48], t = 6.27, p < .001$). No interaction effect was found ($p = .73$) ($R^2 = .179, R^2_{adjusted} = .159$). In the ER, only the main effect of Session was significant, indicating enhanced

Table 4. Performance on the TEA subtests by sessions. SDs are given in parentheses

	L2-instruction group				L2-immersion group			
	W0	W3	W6	W10	W0	W3	W6	W10
EC	99.27 (3.19)	97.8 (7.71)	97.8 (6.16)	98.46 (4.5)	99.25 (3.23)	97.74 (7.07)	100 (0)	96.83 (10.31)
ED	77.69 (20.76)	87.44 (19.27)	95.38 (17.09)	94.59 (15.64)	70 (24.58)	78.16 (27.54)	89.47 (21.52)	91.94 (24.64)
ER	68.21 (21.08)	78.46 (14.46)	79.74 (6.82)	83.24 (8.36)	68.95 (26.61)	73.42 (25.77)	82.63 (18.59)	86.11 (18.02)

**Figure 5.** Performance in the respective TEA Elevator subtests (i.e., Elevator with Distraction-ED and Elevator with Reversal-ER) by group. Error bars represent $\pm 1SE$.

performance in attentional switching with testing session from W3 (W3 vs. W0: $Est = 7.4$, 95% CI [0.55, 14.25], $t = 2.13$, $p = .034$; W6 vs. W0: $Est = 12.6$, 95% CI [5.75, 19.45], $t = 3.62$, $p < .001$; W10 vs. W0: $Est = 16.09$, 95% CI [9.14, 23.03], $t = 4.56$, $p < .001$). No other main effect or interaction effect were significant (all $ps > .05$) ($R^2 = .079$, $R^2_{adjusted} = .057$) (see Figure 5).

Corsi Tapping Task

Performance on each sub-task of the Corsi Tapping Task is given in Table 5.

In the forward condition, no main effects or interaction effects were found (all $ps > .05$) ($R^2 = .028$, $R^2_{adjusted} = .005$). The average accuracy rate was 53.19%. In the backward condition, the main effect of Session was significant: enhanced performance in working memory with testing sessions starting from W6 (W6 vs. W0: $Est = 3.41$, 95% CI [0.38, 6.44], $t = 2.21$, $p = .028$; W10 vs. W0: $Est = 4.17$, 95% CI [1.1, 7.24], $t = 2.67$, $p = .008$) ($R^2 = .039$, $R^2_{adjusted} = .016$). No other main effect or interaction effect were significant (all $ps > .05$). In the rotated condition, the main effect of Session was significant: enhanced performance in mental rotation with testing sessions starting from W3 (W3 vs. W0: $Est = 5.85$, 95% CI [1.2, 10.5], $t = 2.47$, $p = 0.014$; W6 vs. W0: $Est = 8.64$, 95% CI [3.98, 13.29], $t = 3.65$, $p < .001$; W10 vs. W0: $Est = 10.74$, 95% CI [6.02, 15.45], $t = 4.48$, $p < .001$) ($R^2 = .082$, $R^2_{adjusted} = .06$). No other main effect or interaction effect were significant (all $ps > .05$) (see Figure 6).

Table 5. Performance on the Corsi Tapping task by Sessions. SDs are given in parentheses

	L2-instruction group				L2-immersion group			
	W0	W3	W6	W10	W0	W3	W6	W10
Forward	51.76	51.6	53.04	54.05	50.66	55.1	54.11	55.38
	(9.4)	(7.68)	(10.23)	(10.01)	(10.66)	(9.95)	(9.78)	(9.7)
Backward	39.9	42.95	44.87	45.78	43.59	43.26	45.39	46.01
	(9.58)	(9.74)	(8.94)	(11.6)	(8.15)	(10.11)	(8.56)	(9.58)
Rotated	38.03	43.04	44.58	45.27	36.59	43.3	47.37	50.93
	(14.66)	(12.76)	(14.98)	(12.38)	(12.49)	(15.01)	(16.51)	(17.86)

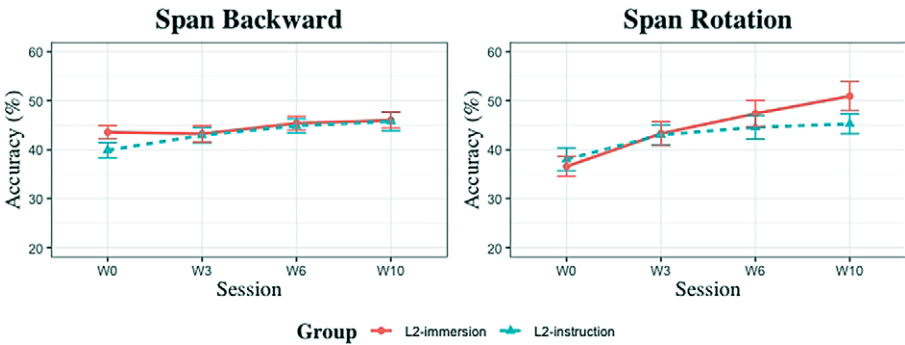


Figure 6. Performance in the respective Corsi Tapping Task (i.e., backward and rotation conditions) by group. Error bars represent $\pm 1SE$.

Discussion

The aim of this study was to longitudinally investigate the effect of different pathways of language acquisition (i.e., instruction vs. immersion) and linguistic context (i.e., L2-dominant linguistic environment) on cognitive functions in young adult Chinese speakers learning English, who moved from China and continued their L2 learning in the United Kingdom. The two groups of participants involved in the current study mainly differed in terms of language acquisition and were comparable on background measures (e.g., demographic and social backgrounds; nonverbal intelligence and SES), life experience-based variables (e.g., musical and video-gaming experience), as well as language-related variables (e.g., AoA and exposure to L2-dominant linguistic environment). A four-repeated-assessment longitudinal design was used to track the possible changes on cognitive functions over 10 weeks. Cognitive performance was assessed through four nonlinguistic cognitive tasks: the ANT, the Number Stroop task, the Elevator subtests of the TEA (EC, ED, and ER), and the CTT, providing multiple-dimensional measures of cognitive functions in both the visual and auditory domains.

Both groups showed similar performance in all cognitive measures at the first testing session, suggesting comparable cognitive performance at the baseline (i.e., W0). Analyses to trace the degree of change over 10 weeks showed that these changes occurred at the earlier stage of testing session and lasted until the last session. The longitudinal results demonstrate a similar enhancement of cognitive performance in both groups in tasks assessing visual and auditory inhibition, attentional switching, and mental rotation, indicating an overall effect of linguistic context. The only significant group difference was observed in the auditory inhibition test, in which the L2-instruction

group showed an overall superior performance than the L2-immersion group, suggesting a specific cognitive effect of language experience.

Specifically, both groups showed a comparable enhancement of inhibitory control in both visual and auditory domains over 10 weeks, suggesting a general effect of linguistic context. Given that both groups moved from their L1 country to the L2 English country, the change of linguistic environment (i.e., from L1-dominant to L2-dominant) might account for these enhancements. According to the Adaptive Control Hypothesis (Green & Abutalebi, 2013), L2 learners might require more interference control in a single language context. In the current study, the Chinese-English participants were living in an L2-dominant linguistic environment (i.e., a single language context), which may lead to enhanced inhibitory control. Specifically, to adjust to the L2 environment, our participants had to actively use their less-dominant L2 while inhibiting their dominant L1 to the furthest extent, especially at the early stage. When they became used to the L2-dominant environment, the increasing demands of related cognitive control might not be as significant as the beginning, which may explain the absence of changes between the testing sessions W3–W6 and W6–W10. Moreover, it has been suggested that the inhibition is greater when inhibiting the dominant L1 than when inhibiting the less-dominant L2 (Costa & Santesteban, 2004). Hence, more cognitive processes were required for inhibiting the dominant L1 in our participants.

The only group difference was observed in auditory inhibition, where the L2-instruction group showed an overall superior performance than the L2-immersion group. This observation indicates a specific effect of language experience on cognitive performance. Previous research has shown that intensive L2 input and exposure predict better L2 competence (Collins *et al.*, 1999; Freed *et al.*, 2004; Serrano *et al.*, 2011) as well as a significant improvement in an attentional switching task (i.e., auditory attentional switching; Bak *et al.*, 2016). Elmer *et al.* (2011), for example, found that intensive language training can influence brain activity in regions involved in auditory attention using the event-related functional magnetic resonance imaging (fMRI). The L2-instruction group in the current study, who attended the PSE program, had longer class hours and received relatively more intensive language training than the L2-immersion group. This specific language experience in the L2-instruction group could lead to the observed superior performance in auditory inhibition than the L2-immersion group. This greater improvement in performance on auditory inhibition could be argued to be the result of greater inhibition effort for L1 in the instruction group given its lower L2 proficiency at W0 compared to the immersion group. It should be noticed that both groups showed comparable improvement in language proficiency over 10 weeks and similar performance in all cognitive measures at the first testing session even though the L2-instruction group had a relatively lower level of proficiency than the L2-immersion group at the baseline. This suggests that language proficiency is not a strong predictor of cognitive performance in the current study.

The effect of language experience on inhibitory control was limited to the auditory domain, but not the visual domain, which could be due to the characteristics of the selected tasks. As Ooi *et al.* (2018) explained that different tasks might differ in assessing attentional control: the ANT and Stroop task were motivated by a theoretical framework of attention and mostly used in an experimental setting, while the Elevator subtests of the TEA were ecologically valid because arguably closer to everyday activities. In the other words, the subtests of the TEA are relatively sensitive in capturing the potential group differences in cognitive behavior measurements in young adults. Moreover, these tasks tapped to different domains: the ANT and Stroop task measured the visual attention control, while the TEA measured the auditory attention control.

Participants in the current study learned English mainly through the auditory modality (i.e., listening and speaking), which is reflected in the time of the L2 learning and usage. These might explain the fact that the L2-instruction group obtained an overall superior performance than the L2-immersion group for inhibitory control measured by the ED subtest of the TEA, but not by the ANT or Stroop task.

In addition to inhibition, a similar improvement of attentional switching abilities was observed in both groups. The results of the time spent on practicing L2 showed that our participants mainly used L2 in classes and the average hours were less than 8 hours per week. This meant that our participants might have frequently switched between their two languages after classes (e.g., communicating with their friends in their L1), which might have led to enhanced attentional switching abilities. Previous studies suggested that bilingualism could increase flexibility in mental-set shifting (Houtzager et al., 2017; Prior & MacWhinney, 2010), which received supporting evidence from neuroimaging studies (Garbin et al., 2010; Gold et al., 2013; Green & Abutalebi, 2013; Rodríguez-Pujadas et al., 2013).

There is existing evidence that an increase in one type of cognitive abilities might positively affect other related abilities (Green & Bavelier, 2003). This could partly explain the improved performance on mental rotation. Mental rotation refers to the cognitive ability to process visual-spatial representations. McLeay (2003), for example, investigated the link between bilingualism and spatial ability and reported that bilinguals showed superior performance in spatial tasks involving mental manipulation compared to monolinguals. Studies have suggested that bilingualism may enhance the ability to control selective attention to specific aspects of mental representations (Bialystok, 1999; Bialystok & Martin, 2004). More specifically, bilinguals may rely more heavily on visual or spatial strategies than their monolingual counterparts, leading to better performance on tasks involving mental imagery and spatial manipulation (Ransdell & Fischler, 1991). The enhancement of L2 proficiency and intensive exposure to L2 in the current study could lead to more efficient performance in tasks involving mental rotation over 10 weeks.

A significant decrease of overall response time over testing sessions was found in both groups in the Stroop task, but not in the ANT. This could be explained as a global RT effect associated with bilingualism as proposed by Hilchey and Klein (2011): a domain-general enhancement of executive control in bilinguals could predict an overall faster response. The overall RTs in the ANT was stable, which could be due to the demand level of the task. The relatively less-demanding ANT, as compared to the Stroop task, might not be sufficient enough to exhibit any change (Qu et al., 2016). The alternative explanation could be the practice effect due to familiarity with the task processes and better performance with repeated testing (Donovan & Radosevich, 1999). However, this could not explain the stable overall RTs in the ANT. Notably, the average accuracy rate did not change in both tasks.

In contrast, stable performance was found in tasks measuring other aspects of attentional control (i.e., alerting and orienting in the ANT and sustained attention in the EC) and working memory (i.e., forward condition in the CTT). Specifically, all participants showed a ceiling effect in the EC and obtained an average accuracy rate around 50% in the forward condition of the CTT. In contrast to studies by Nicolay and Poncelet (2013, 2015), in which significant enhancement on alerting was reported in the immersed children compared to the nonimmersed children. It has been explained that the immersed children had low proficiency in the L2, and thus had to maintain their continuous readiness for effortful processing, such as for monitoring and understanding the academic courses taught in L2. Such experience has led to an enhancement

in alerting abilities by the immersed children. It should be noted that the studies by Nicolay and Poncet (2013, 2015) involved young children, where the potential differences on cognitive performance are relatively “easier” to capture. The behavioral cognitive measures might not be sufficient enough to capture the differences in the current study with young adults, who may perform at ceiling level (Bialystok *et al.*, 2005).

There were no significant interactions among the fixed variables, but the figures seem to suggest a different speed of changes between the two groups on specific cognitive measures, such as conflict resolution in the ANT, auditory attentional performance in the subtests of TEA (*i.e.*, ED and ER), and working memory in the relatively demanding backward condition in the CTT. The reason behind the nonsignificant interactions could be the time interval between the repeated assessments: a longer time interval might not capture the relatively earlier changes between the two groups. In the current longitudinal design, which was based on the timeline of the 10-week PSE, there were four repeated measures (*i.e.*, W0, W3, W6, and W10) and three times of intervals (*i.e.*, 3 weeks, 3 weeks, and 4 weeks). Our results observed initial enhancement in cognitive measures between W0 and W3, but no enhancements between W3–W6 and W6–W10, indicating significantly initial improvement induced by the early-stage testing session. It is possible that there were group differences within Week 1 and the two groups reached comparable performance at Week 3, leading to a statistically nonsignificant interaction. This possible explanation receives supporting evidence from previous studies using the same Elevator subtests of the TEA (Bak *et al.*, 2016; Long *et al.*, 2020), in which language learners showed comparable performance at baseline and displayed significant improvements in the ER at the end of 1-week intensive Gaelic course. Hence, in the future studies, the time interval between repeated assessments could be shorter within a longitudinal research design, such as W0, W1, W2, and W3.

The main limitation in the current study would be the existence of practice and familiarity effects for some tasks, which are well-recognized in psychology (Donovan & Radosevich, 1999). It might be argued that the changes over repeated testing could be due to practice effects, interpreted as automatization of performance (Raichle *et al.*, 1994). In the ANT, there is little evidence that there are major practice effects (Fan *et al.*, 2002). In the current study, no practice effect was observed, as reflected by the stability of overall RTs and accuracy rate over the repeated testing. Furthermore, decreased RTs were only observed on the incongruent trials that involved conflict resolutions, not on the congruent trials, suggesting that the main changes observed in the inhibitory control measured by the ANT for both groups resulted from the L2-dominant linguistic context. In the TEA, three parallel versions were developed to avoid practice effects (Robertson *et al.*, 1994), as the longitudinal follow-up, for example, to evaluate neurorehabilitation, was an important part of the test design. The absence of practice effects in the TEA was confirmed in studies using the TEA to examine the impacts of intensive language learning on attentional functions reported no practice effects (Bak *et al.*, 2016; Long *et al.*, 2020). Hence, changes in performance in our participants are more likely to be due to the language experience and L2-dominant linguistic context than to practice effects. However, no studies have examined practice effects in the Stroop and CTT, the performance assessed by the two tasks could not rule out the potential influence of practice effects. Importantly, practice effects cannot explain the observed difference (*i.e.*, auditory inhibitory control) between the two groups, as we would expect them to be similar in both conditions. Likewise, differences between the

groups are unlikely to have resulted from nonlinguistic effects of the experience of moving to another country, and experience which both groups have shared.

In conclusion, this study presents the first longitudinal design with four repeated assessments aimed at investigating the effects of L2 instruction versus immersion on cognitive functions whilst tracing the observed changes within young adult individuals, whose cognitive capacities are supposed to be at their peak levels. The results suggest a comparable enhancement in most executive functions (i.e., visual and auditory inhibition, attentional switching, and mental rotation) in both groups, but the L2-instruction group showed overall superior performance in auditory attention than the L2-immersion group. These observations suggest a general effect of linguistic context (i.e., L2-dominant linguistic environment) and a specific effect of language experience. Further research could employ a longitudinal design to focus on how specific factors related to the L2 acquisition and bilingual experience affect executive functions within individuals, instead of traditional between-group design (i.e., immersion vs. nonimmersion). Moreover, the linguistic context plays an important role in language acquisition and language control.

Acknowledgments. This work received support of School of Philosophy, Psychology & Language Sciences Research Support Grants.

Supplementary Materials. To view supplementary material for this article, please visit <http://doi.org/10.1017/S0272263122000158>.

References

- Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders*, 36, 233–240. [https://doi.org/10.1016/S0021-9924\(03\)00022-4](https://doi.org/10.1016/S0021-9924(03)00022-4)
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bak, T. H., Long, M. R., Vega-Mendoza, M., & Sorace, A. (2016). Novelty, challenge, and practice: The impact of intensive language learning on attentional functions. *PLoS ONE*, 11. <https://doi.org/10.1371/journal.pone.0153485>
- Barton, K. (2014). MuMIn: Multi-model inference. R package version 1.10.5.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67. <https://doi.org/10.18637/jss.v067.i01>
- Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development*, 70, 636–644. <https://doi.org/10.1111/1467-8624.00046>
- Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, 122, 67–73. <https://doi.org/10.1016/j.cognition.2011.08.003>
- Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental science*, 7, 325–339. <https://doi.org/10.1111/j.1467-7687.2004.00351.x>
- Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics*, 21, 522–538. <https://doi.org/10.1016/j.jneuroling.2007.07.001>
- Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, 9, 103–119. <https://doi.org/10.1177/13670069050090010701>
- Bice, K., & Kroll, J. F. (2019). English only? Monolinguals in linguistically diverse contexts have an edge in language learning. *Brain and Language*, 196, 104644. <https://doi.org/10.1016/j.bandl.2019.104644>
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, 11, 282–298. <https://doi.org/10.1111/j.1467-7687.2008.00675.x>

- Collentine, J. (2008). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition*, 26, 227–248. <https://doi.org/10.1017/S0272263104262040>
- Collins, L., Halter, R. H., Lightbown, P. M., & Spada, N. (1999). Time and the distribution of time in L2 instruction. *TESOL Quarterly*, 33, 655. <https://doi.org/10.2307/3587881>
- Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, 106, 59–86. <https://doi.org/10.1016/j.cognition.2006.12.013>
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50, 491–511. <https://doi.org/10.1016/j.jml.2004.02.002>
- Cummins, J. (1998). Immersion education for the millennium: What have we learned from 30 years of research on second language immersion? In M. R. Childs & R. M. Bostwick (Eds.) *Learning through two languages: Research and practice. Second Katoh Gakuen International Symposium on Immersion and Bilingual Education*. Katoh Gakuen.
- Cummings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28, 369–382. <https://doi.org/10.1177/0267658312443651>
- Dahl, A., & Vulchanova, M. D. (2014). Naturalistic acquisition in an early language classroom. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00329>
- De Graaff, R., Jan Koopman, G., Anikina, Y., & Westhoff, G. (2007). An observation tool for effective L2 pedagogy in Content and Language Integrated Learning (CLIL). *International Journal of Bilingual Education and Bilingualism*, 10, 603–624. <https://doi.org/10.2167/beb462.0>
- DeKeyser, R. M. (2007). Study abroad as foreign language practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.
- Deluca, V., Rothman, J., & Pliatsikas, C. (2019). Linguistic immersion and structural effects on the bilingual brain: A longitudinal study. *Bilingualism: Language and Cognition*, 22, 1160–1175. <https://doi.org/10.1017/S1366728918000883>
- Dewey, D. P. (2004). A comparison of reading development by learners of Japanese in intensive domestic immersion and study abroad contexts. *Studies in Second Language Acquisition*, 26, 303–327. <https://doi.org/10.1017/S0272263104262076>
- Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science*, 21, 335–341. <https://doi.org/10.1177/0963721412453722>
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science*, 333, 959–964. <https://doi.org/10.1126/science.1204529>
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. *Journal of Applied Psychology*, 84, 795–805. <https://doi.org/10.1037/0021-9010.84.5.795>
- Elmer, S., Meyer, M., Marrama, L., & Jäncke, L. (2011). Intensive language training and attention modulate the involvement of fronto-parietal regions during a non-verbal auditory discrimination task. *European Journal of Neuroscience*, 34, 165–175. <https://doi.org/10.1111/j.1460-9568.2011.07728.x>
- Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, 14, 340–347. <https://doi.org/10.1162/089892902317361886>
- 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, 275–301. <https://doi.org/10.1017/S0272263104262064>
- Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., Hernandez, M., Costa, A., & Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *NeuroImage*, 53, 1272–1278. <https://doi.org/10.1016/j.neuroimage.2010.05.078>
- Genesee, F. (2006). What do we know about bilingual education for majority-language students? In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 547–576). Wiley-Blackwell.
- Gold, B. T., Kim, C., Johnson, N. F., Kryscio, R. J., & Smith, C. D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience*, 33, 387–396. <https://doi.org/10.1523/JNEUROSCI.3837-12.2013>
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423, 534–537. <https://doi.org/10.1038/nature01647>

- Green, D. W., & Abutaleb, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 25, 515–530. <https://doi.org/10.1080/20445911.2013.796377>
- Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive Development*, 28, 41–50. <https://doi.org/10.1016/j.cogdev.2012.10.002>
- Grundy, J. G., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, 33, 325–340. <https://doi.org/10.1177/0267658316678286>
- Hernández, M., Costa, A., Fuentes, L. J., Vivas, A. B., & Sebastián-Gallés, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition*, 13, 315–325. <https://doi.org/10.1017/S1366728909990010>
- Hernández, T. A. (2010). The relationship among motivation, interaction, and the development of second language oral proficiency in a study-abroad context. *The Modern Language Journal*, 94, 600–617. <https://doi.org/10.1111/j.1540-4781.2010.01053.x>
- Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, 18, 625–658. <https://doi.org/10.3758/s13423-011-0116-7>
- Houtzager, N., Lowie, W., Sprenger, S., & De Bot, K. (2017). A bilingual advantage in task switching? Age-related differences between German monolinguals and Dutch-Frisian bilinguals. *Bilingualism: Language and Cognition*, 20, 69–79. <https://doi.org/10.1017/S1366728915000498>
- Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology*, 144, 84–97. <https://doi.org/10.1016/j.jecp.2015.11.009>
- Kaushanskaya, M., Gross, M., & Buac, M. (2014). Effects of classroom bilingualism on task-shifting, verbal memory, and word learning in children. *Developmental Science*, 17, 564–583. <https://doi.org/10.1111/desc.12142>
- Keehner, M., & Gathercole, S. E. (2007). Cognitive adaptations arising from nonnative. *Memory & Cognition*, 35, 752–761. <https://doi.org/10.3758/BF03193312>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82. <https://doi.org/10.18637/jss.v082.i13>
- Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language. *Psychological Science*, 20, 1507–1515. <https://doi.org/10.1111/j.1467-9280.2009.02480.x>
- Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, 37, 353–365. <https://doi.org/10.1016/j.system.2009.03.001>
- Long, M. R., Vega-Mendoza, M., Rohde, H., Sorace, A., & Bak, T. H. (2020). Understudied factors contributing to variability in cognitive performance related to language learning. *Bilingualism: Language and Cognition*, 23, 801–811. <https://doi.org/10.1017/S1366728919000749>
- Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25, 605–621. <https://doi.org/10.1080/20445911.2013.795574>
- Malt, B. C., & Sloman, S. A. (2003). Linguistic diversity and object naming by non-native speakers of English. *Bilingualism: Language and Cognition*, 6, 47–67. <https://doi.org/10.1017/S1366728903001020>
- McLeay, H. (2003). The relationship between bilingualism and the performance of spatial tasks. *International Journal of Bilingual Education and Bilingualism*, 6, 423–438. <https://doi.org/10.1080/13670050308667795>
- Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, 21, 8–14. <https://doi.org/10.1177/0963721411429458>
- Nicolay, A.-C., & Poncelet, M. (2013). Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism: Language and Cognition*, 16, 597–607. <https://doi.org/10.1017/S1366728912000375>
- Nicolay, A.-C., & Poncelet, M. (2015). Cognitive benefits in children enrolled in an early bilingual immersion school: A follow up study. *Bilingualism: Language and Cognition*, 18, 789–795. <https://doi.org/10.1017/S1366728914000868>
- Ooi, S. H., Goh, W. D., Sorace, A., & Bak, T. H. (2018). From bilingualism to bilingualisms: Bilingual experience in Edinburgh and Singapore affects attentional control differently. *Bilingualism: Language and Cognition*, 21, 867–879. <https://doi.org/10.1017/S1366728918000020>

- Parafita Couto, M. C., Mueller Gathercole, V. C., & Stadthagen-Gonzalez, H. (2015). Interface strategies in monolingual and end-state L2 Spanish grammars are not that different. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01525>
- Pliatsikas, C., & Chondrogianni, V. (2015). Editorial: Learning a non-native language in a naturalistic environment: insights from behavioral and neuroimaging research. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01009>
- Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, 113, 535–551. <https://doi.org/10.1016/j.jecp.2012.06.013>
- Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, 13, 253–262. <https://doi.org/10.1017/S1366728909990526>
- Qu, L., Low, J. J. W., Zhang, T., Li, H., & Zelazo, P. D. (2016). Bilingual advantage in executive control when task demands are considered. *Bilingualism: Language and Cognition*, 19, 277–293. <https://doi.org/10.1017/S1366728914000376>
- Raichle, M. E., Fiez, J. A., Videz, T. O., MacLeod, A. M. K., Pardo, J. V., Fox, P. T., & Petersen, S. E. (1994). Practice-related changes in human brain functional anatomy during nonmotor learning. *Cerebral Cortex*, 4, 8–26. <https://doi.org/10.1093/cercor/4.1.8>
- Ransdell, S. E., & Fischler, I. (1991). Imagery skill and preference in bilinguals. *Applied Cognitive Psychology*, 5, 97–112. <https://doi.org/10.1002/acp.2350050204>
- Raven, J. C., & Foulds, G. A. (1962). *Advanced Progressive Matrices*. H.K. Lewis & Co. Ltd.
- R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>
- Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). The test of everyday attention (TEA). Thames Valley Test Company, 197–221.
- Rodríguez-Pujadas, A., Sanjuán, A., Ventura-Campos, N., Román, P., Martín, C., Barceló, F., Costa, A., & Ávila, C. (2013). Bilinguals use language-control brain areas more than monolinguals to perform non-linguistic switching tasks. *PLoS ONE*, 8, e73028. <https://doi.org/10.1371/journal.pone.0073028>
- Schaie, K. W. (1996). *Intellectual development in adulthood: The Seattle longitudinal study*. Cambridge University Press.
- Segalowitz, N., Freed, B., Collentine, J., Lafford, B., Lazar, N., & Díaz-Campos, M. (2004). A comparison of Spanish second language acquisition in two different learning contexts: Study abroad and the domestic classroom. *The Interdisciplinary Journal of Study Abroad*, 10, 1–18. <http://hdl.handle.net/2022/25417>
- Serrano, R. (2011). The time factor in EFL classroom practice: Time and EFL practice. *Language Learning*, 61, 117–145. <https://doi.org/10.1111/j.1467-9922.2010.00591.x>
- Serrano, R., Llanes, À., & Tragant, E. (2011). Analyzing the effect of context of second language learning: Domestic intensive and semi-intensive courses vs. study abroad in Europe. *System*, 39, 133–143. <https://doi.org/10.1016/j.system.2011.05.002>
- Simonis, M., Van der Linden, L., Galand, B., Hilgsmann, P., & Szmalec, A. (2020). Executive control performance and foreign-language proficiency associated with immersion education in French-speaking Belgium. *Bilingualism: Language and Cognition*, 23, 355–370. <https://doi.org/10.1017/S136672891900021X>
- Spronken, M., Holland, R. W., Figner, B., & Dijksterhuis, A. (2016). Temporal focus, temporal distance, and mind-wandering valence: Results from an experience sampling and an experimental study. *Consciousness and Cognition*, 41, 104–118. <https://doi.org/10.1016/j.concog.2016.02.004>
- Steinberg, L., Icenogle, G., Shulman, E. P., Breiner, K., Chein, J., Bacchini, D., Chang, L., Chaudhary, N., Giunta, L. D., Dodge, K. A., Fanti, K. A., Lansford, J. E., Malone, P. S., Oburu, P., Pastorelli, C., Skinner, A. T., Sorbring, E., Tapanya, S., Tirado, L. M. U., ... Takash, H. M. S. (2018). Around the world, adolescence is a time of heightened sensation seeking and immature self-regulation. *Developmental Science*, 21, e12532. <https://doi.org/10.1111/desc.12532>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662. <https://doi.org/10.1037/h0054651>
- Tanaka, K., & Ellis, R. (2003). Study-abroad, language proficiency, and learner beliefs about language learning. *JALT Journal*, 25, 63–85. <http://www.jalt-publications.org/archive/jj/2003a/art3.pdf>

- von Bastian, C. C., Souza, A. S., & Gade, M. (2016). No evidence for bilingual cognitive advantages: A test of four hypotheses. *Journal of Experimental Psychology: General*, 145, 246–258. <https://doi.org/10.1037/xge0000120>
- Vulchanova, M., Aurstad, L. M. G., Kvitnes, I. E. N., & Eshuis, H. (2015). As naturalistic as it gets: Subtitles in the English classroom in Norway. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01510>
- Wechsler, D. (1997). Wechsler Memory Scale (WMS-III) (Vol. 14). Psychological Corporation.
- White, J., & Turner, C. (2005). Comparing children's oral ability in two ESL programs. *Canadian Modern Language Review*, 61, 491–517. <https://doi.org/10.3138/cmlr.61.4.491>
- Woumans, E., Surmont, J., Struys, E., & Duyck, W. (2016). The longitudinal effect of bilingual immersion schooling on cognitive control and intelligence: Bilingualism makes you smarter. *Language Learning*, 66, 76–91. <https://doi.org/10.1111/lang.12171>
- Xia, L., Bak, T., Sorace, A., & Vega-Mendoza, M. (2022). Interference suppression in bilingualism: Stimulus-stimulus vs. stimulus-response conflict. *Bilingualism: Language and Cognition*, 25, 256–268. <https://doi.org/10.1017/S1366728921000304>
- Xie, Z., & Pisano, T. S. (2018). Second language (L2) proficiency, socioeconomic status (SES), and intelligence (IQ) are significant predictors of cognitive control differences among young adult unbalanced Chinese–English bilinguals. *Bilingualism: Language and Cognition*, 2000, 1–17. <https://doi.org/10.1017/S1366728918000822>
- Zinszer, B. D., Malt, B. C., Ameer, E., & Li, P. (2014). Native-likeness in second language lexical categorization reflects individual language history and linguistic community norms. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01203>

Cite this article: Xia, L., Bak, T. H., Vega-Mendoza, M. and Sorace, A. (2023). A longitudinal investigation of the effects of language instruction versus immersion on cognitive functions in young adult Chinese speakers learning English. *Studies in Second Language Acquisition*, 45, 189–211. <https://doi.org/10.1017/S0272263122000158>