Studies in Second Language Acquisition, **43** (2021), 329–354
doi:10.1017/S0272263120000534

Research Article

THE EFFECT OF LEARNING CONTEXT ON L2 LISTENING DEVELOPMENT

KNOWLEDGE AND PROCESSING

Xiaoru Yu *

Radboud University Nijmegen International Max Planck Research School

Esther Janse

Radboud University Nijmegen

Rob Schoonen

Radboud University Nijmegen

Abstract
Little research has been done on the effect of learning context on L2 listening development. Motivated by DeKeyser's (2015) skill acquisition theory of second language acquisition, this study compares L2 listening development in study abroad (SA) and at home (AH) contexts from both language knowledge and processing perspectives. One hundred forty-nine Chinese postgraduates studying in either China or the United Kingdom participated in a battery of listening tasks at the beginning and at the end of an academic year. These tasks measure auditory vocabulary knowledge and listening processing efficiency (i.e., accuracy, speed, and stability of processing) in word recognition, grammatical processing, and semantic analysis. Results show that, provided equal starting levels, the SA learners made more progress than the AH learners in speed of processing across the language processing tasks, with less clear results for vocabulary acquisition. Studying abroad may be an effective intervention for L2 learning, especially in terms of processing speed.

This research was partially funded by Chinese Scholarship Council and COST Action CA15130. We thank Prof. Jean-Marc Dewaele and Prof. Dongmei Ma for assistance with data collection and our participants for their interest and contribution.

* Correspondence concerning this article should be addressed to Xiaoru Yu, Centre for Language Studies, Department of Language and Communication, Radboud University Nijmegen, Erasmusplein 1, PO Box 9103, 6500 HD Nijmegen. E-mail: X.Yu@let.ru.nl

© The Author(s), 2020. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (http://creativecommons.org/licenses/by-nc-sa/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

INTRODUCTION

Second language (L2) learning contexts vary widely in quality and quantity of input, output, and interaction, inevitably leading to different L2 development patterns and attainments. Cross-context comparisons, for example, comparing study-abroad (SA) contexts and at-home (AH) contexts, may help to reveal the unique characteristics of L2 development in different learning contexts (Kroll et al., 2018). L2 development can be examined from two distinct perspectives, that is, language knowledge (e.g., vocabulary and grammar) and processing skills (e.g., how rapidly or easily one can understand a sentence). As argued by DeKeyser (2007, 2015) and Hulstijn et al. (2009), L2 learners have to accumulate knowledge of the target language, as well as improve the efficiency with which that knowledge can be processed. Previous studies (e.g., Collentine, 2004; Freed et al., 2004; Håkansson & Norrby, 2010; Pliatsikas & Marinis, 2013a, 2013b; Sasaki, 2007; Segalowitz & Freed, 2004) have investigated how learning context affects the acquisition of both knowledge and processing aspects of language proficiency. For instance, Collentine (2004) analyzed speech produced by American learners of Spanish (in SA and regular-classroom settings) in an Oral Proficiency Interview before and after a semester. All participants were university students with no previous contact with Spanish. The results showed that formal education in an AH context facilitated development of discrete grammatical and lexical knowledge (indicated by use of grammatically marked forms and lexical frequencies, respectively), while the immersion in an SA context was beneficial for the development of oral fluency, a form of processing ability in speaking.

However, studies of L2 learning contexts have rarely focused on listening comprehension, an important but often neglected area of second language acquisition. The few listening studies that have investigated L2 learning contexts (e.g., Cubillos et al., 2008; Llanes & Muñoz, 2009) have largely employed holistic measurement methods (e.g., spoken passage comprehension) to test listening proficiency. Consequently, our knowledge of the effect of learning contexts on the finer components of listening comprehension is rather limited. Therefore, it is not known whether SA learners will be able to recognize more words, or whether they will be faster in recognizing a word, processing grammatical information, and forming the semantic proposition of a sentence in listening comprehension than their AH peers. To fill this gap, this study takes a componential view of listening proficiency, in terms of auditory vocabulary knowledge and listening processing efficiency, and examines L2 listening development in SA and AH learning contexts for an academic year. By comparing SA learners with their AH peers, we aim to investigate whether and to what extent a shift of learning context from AH to SA is an effective intervention for improving adult L2 learners' listening proficiency.

STUDY-ABROAD VERSUS AT-HOME LEARNING CONTEXTS

Studying abroad is often considered as the best L2 learning context (Freed, 1995), as it usually involves a language environment shift where learners have to inhibit the first language and immerse themselves in the target language (Jacobs et al., 2016; Linck et al., 2009). Studying in a country where students' L2 is spoken as the native language guarantees abundant native input, opportunities for output, informative feedback, and interaction. Contrastively, even though nowadays it is relatively easy to have access to

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

authentic L2 input through digital platforms (e.g., Netflix, Audible, and Apple News), AH contexts may be criticized for relatively inadequate L2 exposure, overreliance on rote learning, and limited opportunities for interactive communication. The potential advantages of SA contexts over AH contexts argue for the linguistic benefits of studying abroad. In an ideal scenario where learners are fully immersed in their new environment, SA contexts seem to solve the problems leading to generally low attainments of adult L2 learners in AH contexts. However, SA experiences are hardly ever ideal and immersion degrees may sometimes be overestimated. Newcomers usually experience a gradual process of socialization, starting with compatriots, then expanding to other international students, and finally to locals (Coleman, 2013). As communication across linguistic and cultural boundaries is challenging, the socialization process may stagnate at any stage, as international students may tend to foreground their national identity against intercultural identities during the intercultural experience (Maeder-Qian, 2018). It is not rare that same-nationality students clutter in and out of class. Consequently, the linguistic impact of studying abroad may be compromised due to integration problems.

During the past two decades, blooming international education has led to multiple SA studies (e.g., Dwyer, 2004; Leong, 2007; Sasaki, 2011; Segalowitz & Freed, 2004; Williams, 2005). These studies have been set up from various research angles, focusing on issues such as language proficiency, cross-cultural competencies, personality changes, and career growth. Varela (2017) performed a meta-analysis of 33 studies on language development in SA learners, focusing on dependent variables such as general language proficiency, written proficiency, vocabulary size, and speech rate. Varela reported a large effect of studying abroad on enhancing L2 proficiency ($d = 0.975$). However, this meta-analysis was largely limited to comparing pre- and posttest differences in the SA context. By comparing with Plonsky's (2011) meta-analysis on language learning in the AH context ($d = 0.55$), Varela claimed that SA programs facilitated second language acquisition. Note that this comparison is not direct. As for individual studies that directly contrasted SA and AH contexts, some found that SA learners had greater gains in knowledge of nativelike language usage (Foster et al., 2014), use of communication strategies (Lafford, 2004), grammar (Marqués-Pascual, 2011; Pliatsikas & Marinis, 2013b), accent (Martinsen et al., 2014), pragmatic competence (Matsumura, 2001), writing proficiency (Sasaki, 2011), and oral proficiency (Segalowitz & Freed, 2004), but others reported marginal or no differences as a function of learning context in terms of grammar (Isabelli-García, 2010; Pliatsikas & Marinis, 2013a), and pragmatic comprehension (Taguchi, 2011). Mixed outcomes may relate to the fact that different studies focused on different aspects of language acquisition, which may not be equally sensitive to the effect of learning context.

Furthermore, previous studies have investigated the effects of learning context on fluency, accuracy, and complexity of L2 oral and written production. Learners' oral fluency, measured by speech rate and mean run length with no pause, usually improves after studying abroad (Mora & Valls-Ferrer, 2012; Segalowitz and Freed, 2004). However, accuracy and complexity measures of oral production, such as frequency of errors, length, and syntactic complexity of sentences, shows conflicting results across studies. Some studies provided evidence that SA groups show gains in L2 grammatical complexity and accuracy relative to AH groups (Håkansson & Norrby, 2010; Howard, 2001; Llanes & Muñoz, 2013; Marqués-Pascual, 2011). Other studies, however, claimed that

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

learners achieved better fluency only by using appropriate fillers, modifiers, formulae, and compensation strategies, while their grammatical competence remained unchanged (see Collentine, 2004; DeKeyser, 1991). Similarly, for written production, the benefits of studying abroad have usually been reported to manifest on writing fluency but not necessarily on measures of accuracy and complexity (see Knoch et al., 2015; Sasaki, 2007). These results together seem to suggest that learning contexts have differential effects on different aspects of L2 production. Yet it remains unclear how and to what extent studying abroad affects various aspects of L2 listening comprehension. We set out to evaluate the impact of an SA context, in comparison to AH contexts, on L2 listening development in terms of auditory vocabulary knowledge and processing efficiency.

VOCABULARY KNOWLEDGE

As a type of declarative knowledge, L2 vocabulary knowledge can be acquired either incidentally (i.e., through reading and listening activities aimed at communication and not explicitly at vocabulary learning), or intentionally (i.e., through deliberate memorization of lexical information to enlarge vocabulary size of a target language). *Incidental learning* is widely held to be the major source for accumulating vocabulary knowledge in both L1 and L2 learners, whereas only a relatively small amount of vocabulary is acquired through *intentional learning* (Hulstijn, 2003). Studies of incidental learning reported significant vocabulary gains through extensive reading by L2 learners (Horst, 2005; Pellicer-Sánchez & Schmitt, 2010; Swanborn & De Glopper, 2002). However, L2 incidental vocabulary learning has been associated with low retention rates, which is why some studies (e.g., Horst et al., 1998; Waring & Takaki, 2003) have claimed that the role incidental learning plays in L2 vocabulary acquisition may have been overestimated. Intentional learning, however, has been found to be much more effective than incidental learning in retaining lexical information, especially over a short period (e.g., Schmitt, 2008; Swanborn & De Glopper, 1999).

SA contexts may be superior to AH contexts in facilitating vocabulary acquisition for several reasons. Firstly, the naturalistic exposure in SA contexts arguably guarantees more opportunities for incidental vocabulary learning than AH contexts, with the latter likely being explicitly geared to intentional learning by memorization. Secondly, interaction and negotiation of meaning have been found to facilitate L2 vocabulary acquisition (Ellis et al., 1994; Long, 1996; Newton, 2013). Through negotiating meaning (e.g., by rephrasing or asking for clarification), learners and their interlocutors overcome comprehension difficulties, which may then become learning opportunities. Immersion in SA contexts allows learners to interact and negotiate in the target language, while regular classroom settings in AH contexts are often criticized for limited opportunities for interaction. Thirdly, according to Mayer's (2009) Cognitive Theory of Multimedia Learning, the brain integrates information from visual and auditory channels (e.g., words, pictures, and auditory information) to create mental representations. If visual and auditory channels provide congruent information, people learn better from both channels than only from one channel. Learning words solely from reading, therefore, may not be as efficient as in combination with their auditory form. L2 learners may vocalize words acquired through reading in a deviant form, and consequently may not recognize them in the correctly pronounced form in listening activities. Learners from AH contexts often suffer

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

from the lack of auditory form because listening proficiency is not capitalized on in educational settings, and this problem is supposed to be reduced to some degree for SA learners immersed in their L2. The question as to whether SA contexts better facilitate vocabulary acquisition than AH contexts boils down to whether and to what extent L2 learners can benefit from the incidental learning opportunities provided by an SA context.

Previous studies have compared vocabulary acquisition across different learning contexts (DeKeyser, 1991; Dewey, 2008; Ife et al., 2000; Llanes & Muñoz, 2009; Milton & Meara, 1995). Milton and Meara (1995), for instance, compared students' half-yearly vocabulary growth before and after the onset of their 6-month SA program, and found that the average growth rate in an SA context was four times bigger than that in an AH context. However, Dewey (2008) compared vocabulary gains made by intermediate English learners of Japanese in three learning contexts during 9 to 13 weeks with various vocabulary tests. Participants were either in a SA program, in an intensive domestic immersion program, or in a formal classroom setting. Dewey (2008) found that vocabulary gains in the SA context were not significantly different from those in the intensive domestic immersion setting. This study suggested that the benefits of SA contexts on vocabulary acquisition might stem only from the *amount* of language exposure, rather than the difference in learning contexts. Note that most of the previously mentioned studies tested vocabulary knowledge in reading. Our knowledge about the relationship between learning contexts and auditory vocabulary, which is related to but different from reading vocabulary, is rather limited. Therefore, this study compares auditory vocabulary acquisition in SA, AH-regular classroom, and AH-intensive instruction settings to examine the relationship between auditory vocabulary acquisition and learning context.

PROCESSING EFFICIENCY

According to DeKeyser's (2015) skill acquisition theory of second language acquisition, learners go through declarative, procedural, and automatic stages sequentially during the acquisition process. Firstly, L2 learners start with explicitly learning declarative knowledge (e.g., vocabulary and grammar). Secondly, learners develop procedural knowledge, which is the knowledge exercised in the accomplishment of a task, after a few practice trials. This proceduralization of knowledge is realized when the execution of a target performance gets routinized or chunked (Anderson, 2007; Taatgen & Lee, 2003). This procedural stage is not particularly time-consuming (DeKeyser, 1997). Finally, to use language spontaneously or effortlessly, learners need a large amount of practice to automatize the procedural knowledge acquired in the previous stage. During this slow and gradual process of automatization, learners will comprehend and produce language in a more rapid way, showing fewer errors and requiring less attention. As the automatization of L2 processing is slow (DeKeyser, 2015; Lim & Godfroid, 2015), it is difficult to observe rapid progress in regular foreign language classroom settings. However, the dramatic environmental shift entailed by a SA experience may accelerate L2 automatization, thus creating a situation to test hypotheses about the development of L2 processing skills over a relatively short period. This study describes L2 processing skills during listening comprehension in terms of processing efficiency to avoid the (related, but theoretically charged) term "automaticity" (for a review on automaticity, see Segalowitz,

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

2003, 2010). Processing efficiency is operationalized as a multidimensional construct comprising accuracy, speed, and stability of processing.

Multiple previous studies of second language processing have compared L1 and L2 processing in word recognition, parsing, semantic, or phonological processing (for a review, see Jiang, 2018). Rather than contrasting L1 and L2 processing, only a few studies investigated the development of L2 processing in relation to language learning contexts. For example, Segalowitz and Freed (2004) compared the performance of English (SA and AH) learners of Spanish in a semantic classification task before and after one semester. They reported no effect of learning context on lexical access (quantified as speed of semantic classification decisions). As for studies on grammar acquisition, Isabelli-García (2010) investigated gender acquisition (using grammaticality judgment tests) in intermediate English (SA and AH) learners of Spanish over 4 months, and Pliatsikas and Marinis (2013a) studied the processing of past tense morphology in highly proficient Greek learners of English (one group with more than a year of SA experience and another with only regular AH classroom exposure) with a self-paced reading task. Both studies reported no effect of learning context on grammar acquisition (i.e., gender agreement and past tense, respectively). However, these studies all measure L2 processing capacities in reading, an activity that is emphasized and well-practiced in AH learning contexts. It is not clear how the various cognitive processing abilities in L2 listening comprehension develop across learning contexts. This study investigates the effect of learning contexts on processing abilities at three different levels of listening comprehension, that is, lexical, morphosyntactic, and semantic levels. More specifically, a series of tasks have been devised to measure L2 processing efficiency in lexical access (e.g., recognizing a spoken word), grammatical processing (e.g., capturing a grammatical feature of an utterance), and semantic processing (e.g., understanding the semantic meaning of an utterance). These three processes are critical building blocks toward successful language comprehension (for language comprehension models, see Anderson, 2015; Cutler and Clifton, 1999; Goss, 1982).

CURRENT STUDY

We hypothesize that advanced English learners who study abroad and have experienced a shift of language environment from an English as a Foreign Language (EFL) country (China) to an English as a Native Language (ENL) country (the United Kingdom) will make more progress than their domestic counterparts in terms of both vocabulary size and language processing efficiency. To test this hypothesis, we invited Chinese international non-English-major postgraduates studying in the United Kingdom (SA group, with no previous SA experience), Chinese domestic English-major postgraduates (AH-intensive group) and domestic non-English-major postgraduates (AH-regular group) to participate in a series of English tests at the beginning of their postgraduate program and again after one academic year.

With respect to baseline proficiency, the SA group can be expected to be similar to the AH-regular group because both groups were majoring in non-English subjects. At the same time, the SA group had done intensive preparation, including intensive English learning, to qualify for studying abroad. Consequently, their baseline proficiency may also turn out to be more similar to that of the AH-intensive group. As baseline language

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

proficiency has been shown to relate to size of language learning gains over time (Brecht & Robinson, 1995; Davidson, 2010), both AH groups were included as reference groups for the SA group, to provide a more complete comparison between SA and AH learning contexts.

The pretest and posttest design allows us to compare L2 listening proficiency improvement, in terms of processing efficiency and auditory vocabulary size, across different learning contexts. To investigate the effect of learning contexts on L2 listening development, this study sets out to answer the following questions:

1. Does the SA group show more improvement in auditory vocabulary size than the two AH groups over the course of an academic year?
2. Does the SA group show more improvement in language processing efficiency (i.e., accuracy, speed, and stability of processing) than the two AH groups over the course of an academic year? And if so, is the group difference in improvement constrained to specific linguistic abilities (i.e., lexical access, grammatical processing, and semantic processing)?

METHODOLOGY

PARTICIPANTS

One hundred forty-nine Chinese postgraduates studying abroad or domestically took the pretest and posttest with an interval of 7 months. Among them, there were 47 non-English majors studying in the United Kingdom (SA group), 53 non-English majors studying in China (AH-regular group), and 49 English majors studying in China (AH-intensive group). All the participants finished bachelor education in China, with no previous SA experience before the pretest. The SA group is the target group representing a rapidly increasing population of L2 learners who start learning English as an FL at home and later on move to an English-speaking country as an adult to participate in an SA program. The two AH groups are both control groups to be compared with the SA group to compare English proficiency improvement across different learning contexts.

Before our pretest, the English courses of the AH-intensive group included Comprehensive English, Oral English, Listening Comprehension, Intensive Reading, Extensive Reading, Writing, Literature, Linguistics (around 1,620 hours in total prescribed by their bachelor programs). The AH-regular group and SA group only had a College English class once per week (around 144 hours prescribed by their bachelor program). According to their standardized test scores (see Table 1), at some point during their bachelor program, the AH-regular and the SA groups did not differ in L2 proficiency, while the AH-intensive group had significantly higher language proficiency than the AH-regular and the SA group. Afterward, the SA group, nevertheless, was expected to have learned English (mainly out of class) more than the AH-regular group because they had to prepare for studying abroad. Therefore, we estimated that the baseline language proficiency of the SA group *at pretest* may be somewhere between that of the AH-regular (non-English-major) and AH-intensive (English-major) group.

Between our pre- and posttest, the AH-intensive group had English-medium courses (given by Chinese teachers), which included Literature, Linguistics, Translation, Interpretation, and Methodology (around 468 hours in total), but no basic language learning courses. The SA group also had English-medium courses (around 420 hours in total) but no

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

TABLE 1. Background information of these participant groups: AH-regular group (N = 53), AH-intensive group (N = 49), and SA group (N = 47)

Group	Mean age (SD)[1]	Major	Hours of English instruction during bachelor program (before pretest)[2]	Hour of English instruction during master program (between pre- and posttest)[2]	Mean standardized test score (SD)[3]
AH-regular	23.2 (1.0)	Non-English	144 hours	72 hours	542 (32.6)
AH-intensive	23.4 (1.4)	English	1620 hours	468 hours	563 (38.8)
SA	23.3 (1.1)	Non-English	144 hours	420 hours	528 (54.7)

[1]All participants reported their age in the posttest questionnaire, while this information was not complete in the pretest questionnaire. The means and SDs of Age hereof were therefore calculated based on the posttest questionnaire for the sake of data completeness.
[2]Hours of instruction were calculated based on the credits required by each educational program.
[3]Not all nonnative participants reported their CET-6 (i.e., a national standardized English test) score. The means and SDs for this variable are therefore based on smaller sample sizes (i.e., 147 out of 165). An ANOVA test shows that the AH-regular and the SA group did not differ significantly (p = .221), while the AH-intensive group significantly outperformed the AH-regular (p = .03) and the SA (p = .0003) group in CET-6. However, these results have to be interpreted with caution because participants took the CET-6 test up to 3 years before the pretest, and hence maybe well before the SA group started to prepare for studying abroad.

basic language learning course. The AH-regular group had a 2-hour college English class every week (around 72 hours in total). Therefore, though the SA group was not majoring in English, both the SA and the AH-intensive group would have English-medium education in the coming academic year. Contrastively, only the AH-regular group would mainly have Chinese-medium education.

MATERIALS

The testing materials include a lexical access task, a grammatical processing task, a semantic processing task, and the Peabody Picture Vocabulary Test Fourth Edition (PPVT™-4).

The first three tasks were timed decision tasks used to measure language processing efficiency at the lexical, morphosyntactic, and semantic level, respectively. Measures of these tasks were accuracy, reaction time (RT), and coefficient of variation (CV), indicating accuracy, speed, and stability of processing, respectively. These tasks are designed to focus on how efficiently learners can process their L2 (e.g., how rapidly they can recognize a word, how easily they process a certain grammatical structure, or how fast they can understand an utterance), in addition to their accuracy in performing these tasks. We specifically aimed to minimize the effects of limited vocabulary or grammar knowledge when performing language processing tasks. L2 listening tests often involve holistic measurement methods like spoken passage comprehension. Scores thereof are used as a general indicator of listening proficiency, but reflect little about listening effort or about different components of the listening process. Test stimuli in our tasks are carefully manipulated to allow for detecting the subtle nuances in language processing that are not easily identifiable by existing

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

standardized tests. Therefore, the use of these speeded-response tasks will allow us to investigate the effect of learning context on the development of L2 processing skills.

The (untimed) PPVT™-4 was used to measure auditory vocabulary size, a form of declarative knowledge. The measure of this task was a score for the number of vocabulary items correctly identified, functioning as an approximation of auditory vocabulary size. The PPVT is designed to measure receptive (auditory) vocabulary size of English native speakers aged from 2:6 to more than 90 years. The test's vocabulary items are not restricted to the vocabulary of any specific purpose or discipline. This test is not subject to ceiling effects, either. This format of the test (i.e., choosing a picture that matched the word participants had heard) is more intuitive and straightforward than the commonly used multiple-choice format that involves choosing among synonyms or descriptions.

Lexical Access Task

The lexical access task measured how well participants could recognize words with simultaneous presentation of auditory and visual stimuli. In each trial, participants saw a (line drawing) picture and heard a word simultaneously. They were asked to judge as fast as possible whether the picture and word matched or not by pressing correspondent buttons on a button box. If a picture and a word did not match, their referents shared phonetic similarities (e.g., "kite" and "cat"), fell into the same semantic field (e.g., "apple" and "orange"), or were unrelated (e.g., "frog" and "doctor"). Note that a trial would end and the next trial would start automatically if no response was given within 4 seconds.

This test contained 6 training trials and 60 experimental trials. Cronbach's alpha indices for the accuracy measures were .6 (pretest) and .58 (posttest), while those for the RTs were .97 (pretest) and .96 (posttest). The lower alphas for accuracy are due to the (intended) relative ease of the tasks so that enough valid RTs could be collected.

Grammatical Processing Task

The grammatical processing task, including 6 training trials and 60 experimental trials, tapped into the processing of particular grammatical properties. Participants listened to a sentence, saw two pictures simultaneously, and were asked to quickly choose (within 8 seconds) the picture that matched the sentence they had heard by pressing a corresponding button on a button box. To make the correct choice, participants had to capture a grammatical cue of that sentence. These grammatical cues could be put into two categories: morpho-syntactic cues and function words. Morpho-syntactic cues included plural "-s," third-person singular "-s," tense, aspect, dative, passive and cleft constructions, and relative clause. For example, participants heard the sentence "the sheep eats" and saw two pictures. In the first picture, there were three sheep eating, while in the second one there was only one sheep eating. The morpheme "-s" in the sentence was the key information leading to the correct picture. For another example, participants heard the sentence "It is the dog that the pig follows" and saw two pictures (one with a dog following a pig, while in the other one there was a pig following a dog). The cleft sentence structure was the syntactic cue in this case. As for function words, this category of sentences contained word-level cues, for example, prepositions and conjunctions. For example, the participants heard the sentence "The children are marching along the sidewalk" and saw

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

two pictures: In the first one the children were marching across the sidewalk, and in the second one they were marching along the sidewalk. The preposition "along" in the sentence was the function-word cue indicating the correct picture. The function-word category (20 items out of the total of 60 items) was later excluded for analysis so that this task could better qualify as an indicator of grammatical/morpho-syntactic processing.

The sentence and picture stimuli of this task were partly taken from Kersten (2010), Waters et al. (1995), and Weist (2002). We adapted these stimuli into a time-constrained sentence-picture matching test format. Cronbach's alpha indices for the accuracy measures were .57 (pretest) and .55 (posttest), while those for the RTs were .88 (pretest) and .89 (posttest).

Semantic Processing Task

The semantic processing task measured how efficiently participants could form a semantic interpretation of a sentence. Participants were asked to quickly indicate whether a sentence they were listening to was plausible or not. If a sentence was implausible, it violated either obvious factual knowledge (e.g., "A horse is an animal that can fly") or logic (e.g., "If you eat too much, you can get too thin"). The maximum reaction time for each stimulus was set to 8 seconds from audio onset. This task, containing 6 practice trials and 50 experiment trials, originated from Lim and Godfroid (2015). Cronbach's alpha indices for the accuracy measures were .82 (pretest) and .79 (posttest), while those for the RTs were .95 (pretest) and .95 (posttest).

Vocabulary Size Test

The PPVT™-4 (Dunn & Dunn, 2007) was used to measure auditory vocabulary size. Participants heard a word and saw four pictures on a computer screen, and were asked to choose a picture that matched the word they had just heard. A total of 228 test items are grouped into 19 sets of 12 words, which are arranged of increasing difficulty. Test administration (20 minutes on average) ended automatically if participants had made more than eight errors in one set. Unlike the first three tasks, the vocabulary size test is not timed. Participants could listen to a word multiple times if necessary. Due to the adaptiveness of the test, participants were administered different subsets of items, which prevented us from computing Cronbach's alpha. The PPVT has a reported reliability of .97 (Dunn & Dunn, 2007).

PROCEDURES

Data Collection

The location of data collection was Southeast University at Nanjing, China (for the two AH groups) and Birkbeck, University of London and University College London in the United Kingdom (for the SA group). Background questionnaires were sent out beforehand to screen participants for their eligibility. Eligible participants were invited to take the pretest and the posttest at the beginning and end of an academic year, respectively. After each round of data collection, participants received a small financial reward for their participation.

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

Data Cleaning

Firstly, six items from the grammatical processing and three items from the semantic processing task were excluded as ambiguous items. Items were excluded if they elicited accuracy rates below 80% in native-speaker participants of a previous study (Yu et al., 2020). Secondly, two participants from the AH-intensive group were excluded due to either a lower than 50% accuracy rate on any speeded-response test or because their vocabulary score was not within three standard deviations of the group mean. Thirdly, only RTs of valid (i.e., correct) responses were analyzed. In addition, RTs below 250 ms (measured from audio onset) were removed as invalid responses.

Statistical Analysis

Data analysis was conducted in R version 3.5.1 (R Core Team, 2018). The *glmer* and *lmer* functions in the package *lme4* (Bates et al., 2015) were used to fit logistic and linear mixed-effects regression models (LMMs), with the optimizer bobyqa. P values were calculated and added into linear regression model outputs with the package *lmerTest* (Kuznetsova et al., 2017).

Mixed-effects regression models were fitted to predict accuracy (in logit), speed, and stability of responses in the language processing tasks. RTs were log-transformed before model fitting to normalize their distribution. All models took Time, Group, and Task as fixed-effects predictors, and included maximal by-time, by-participant, and by-item random intercepts and slopes whenever applicable on the premise of model convergence. As the audio duration of test stimuli and trial number may affect participants' reaction times, these two factors were also entered in the RT model as fixed-effect control variables. Because CVs were calculated as an aggregated measure on task level (i.e., $CV = SD_{RT}/Mean_{RT}$), the CV model did not have any item-level random-effect variable. Finally, a linear mixed-effects regression model, with Time and Group as fixed-effect factors and by-participant random effects, was fitted to predict vocabulary size. For all the models, we sum-coded Group and Task variables to compare the main effects of these variables. However, the Time variable was dummy-coded, with pretest performance being mapped on the reference level, to examine group and task effects at pretest. Note that the use of mixed-effects regression models allows us to compare between-group differences in terms of their progress over the academic year while statistically controlling for preexisting between-group differences at pretest.

RESULTS

Table 2 displays the descriptive statistics of the task performances of the participant groups in the pre- and posttests. The rest of this section gives a detailed description of the statistical results of the accuracy, RT, CV, and vocabulary size models. For the description of the model results, Group and Task effects at pretest will be presented before effects concerning the Time variable (i.e., the difference between the pretest and the posttest), such as the main effect of Time, two-way interactions between Time and Group, and three-way interactions between Time, Group, and Task. The interactions between Time

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

TABLE 2. Descriptive statistics of task performance in the pre- and posttests for AH non-English majors (AH-regular, $N = 53$), AH English majors (AH-intensive, $N = 49$), and the SA group (SA, $N = 47$)

			Pretest	Posttest
Task	Measure	Group	Mean (*SD*)	Mean (*SD*)
Lexical access task	Accuracy	AH-regular	.83 (.06)	.83 (.05)
		AH-intensive	.87 (.05)	.88 (.05)
		SA	.88 (.04)	.89 (.04)
	RT	AH-regular	1,284 (291)	1,038 (194)
		AH-intensive	1,188 (265)	1,080 (223)
		SA	1,137 (250)	981 (190)
	CV	AH-regular	.36 (.09)	.31 (.09)
		AH-intensive	.34 (.09)	.33 (.08)
		SA	.33 (.08)	.29 (.09)
Grammatical processing task	Accuracy	AH-regular	.83 (.08)	.85 (.08)
		AH-intensive	.86 (.06)	.88 (.06)
		SA	.86 (.07)	.88 (.05)
	RT	AH-regular	3,375 (452)	3,008 (458)
		AH-intensive	3,248 (436)	2,936 (371)
		SA	3,289 (424)	2,871 (379)
	CV	AH-regular	.38 (.05)	.38 (.05)
		AH-intensive	.39 (.06)	.39 (.05)
		SA	.38 (.05)	.39 (.06)
Semantic processing task	Accuracy	AH-regular	.84 (.12)	.87 (.11)
		AH-intensive	.93 (.05)	.95 (.04)
		SA	.93 (.07)	.94 (.05)
	RT	AH-regular	3,596 (669)	3,061 (567)
		AH-intensive	3,011 (347)	2,737 (308)
		SA	3,049 (447)	2,691 (441)
	CV	AH-regular	.29 (.06)	.25 (.05)
		AH-intensive	.28 (.06)	.26 (.05)
		SA	.26 (.05)	.23 (.06)
Vocabulary size test	Vocabulary size score	AH-regular	116 (28)	121 (27)
		AH-intensive	148 (17)	149 (17)
		SA	132 (23)	139 (22)

Note: The values of Accuracy are proportions and RT values are in milliseconds.

and Group, and the interactions between Time, Group, and Task are most critical for answering the research questions of this study.

VOCABULARY SIZE

As Table 3 and Figure 1 show, the AH-intensive group outperformed the SA group ($\beta = 15.97$, $SE = 4.69$, $t = 3.40$, $p < .001$), who in turn outperformed the AH-regular group ($\beta = -15.12$, $SE = 4.58$, $t = -3.30$, $p < .001$) in terms of vocabulary in the *pretest*. In general, participants' vocabulary size in the posttest was significantly larger than in the pretest ($\beta = 4.56$, $SE = 1.11$, $t = 4.10$, $p < .001$). The SA group made more progress than the AH-intensive group ($\beta = -6.09$, $SE = 2.77$, $t = -2.20$, $p = .030$), but there was no

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

TABLE 3. Estimates of performance of participant groups in the vocabulary size test

	B	*SE*	*t*	*P*
(Intercept)	131.86	1.88	70.09	< .001
TimeT2vsT1	4.56	1.11	4.10	< .001
GroupAH-regularvsSA	−15.12	4.58	−3.30	< .001
GroupAH-intensivevsSA	15.97	4.69	3.40	< .001
TimeT2vsT1:GroupAH-regularvsSA	−2.58	2.71	−0.95	0.342
TimeT2vsT1:GroupAH-intensivevsSA	−6.09	2.77	−2.20	0.030

Note: Model specification: lmer(Vocab ~ Time * Group + (1|Subject)).
T1 and T2 refer to the pretest and posttest, respectively.

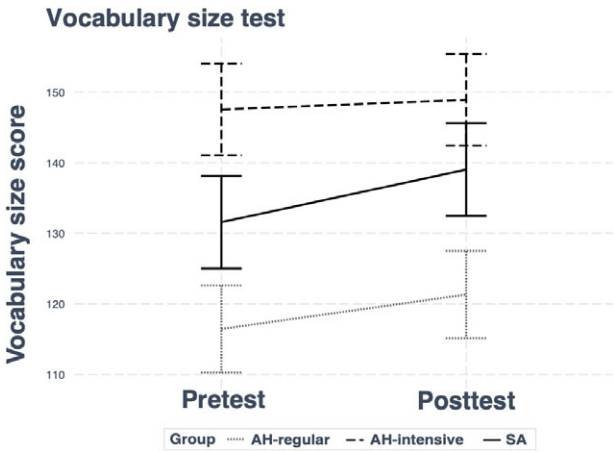

FIGURE 1. Interaction between Time and Group effects in the vocabulary size model. Y-axis does not start from zero. Error bars represent standard errors.

significant difference between vocabulary improvement of the SA and the AH-regular groups ($\beta = -2.58$, $SE = 2.71$, $t = -0.95$, $p = .342$). Furthermore, we split the data by group (see Appendix Table A1) and found that both the SA group ($\beta = 7.45$, $SE = 1.80$, $t = 4.14$, p < .001) and the AH-regular group ($\beta = 4.87$, $SE = 2.29$, $t = 2.13$, $p = .038$) made significant improvement in vocabulary over time but the AH-intensive group ($\beta = 1.35$, $SE = 1.49$, $t = 0.91$, $p = .369$) did not.

PROCESSING ACCURACY

As shown in Table 4 and Figure 2, the SA group outperformed the AH-regular group ($\beta = -0.80$, $SE = 0.13$, $z = -6.37$, $p < .001$), but did not differ from the AH-intensive group ($\beta = -0.16$, $SE = 0.13$, $z = -1.19$, $p = .233$) in terms of processing accuracy in the *pretest*. Accuracy performance in the pretest did not differ across tasks ($\beta = -0.52$, $SE = 0.30$, $z = -1.73$, $p = .084$; $\beta = 0.10$, $SE = 0.31$, $z = 0.32$, $p = .748$). However, the difference

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

TABLE 4. Fixed-effect estimates of accuracy performance of participant groups in the three processing tasks

	B	*SE*	*z*	*p*
(Intercept)	2.74	0.13	20.56	< .001
TimeT2vsT1	0.23	0.04	5.15	< .001
GroupAH-regularvsSA	−0.80	0.13	−6.37	< .001
GroupAH-intensivevsSA	−0.16	0.13	−1.19	.233
Task2vs1	−0.52	0.30	−1.73	.084
Task3vs1	0.10	0.31	0.32	.748
TimeT2vsT1:GroupAH-regularvsSA	−0.09	0.08	−1.11	.269
TimeT2vsT1:GroupAH-intensivevsSA	0.05	0.09	0.58	.565
TimeT2vsT1:Task2vs1	0.15	0.10	1.46	.146
TimeT2vsT1:Task3vs1	0.09	0.11	0.80	.424
GroupAH-regularvsSA:Task2vs1	0.44	0.17	2.55	.011
GroupAH-intensivevsSA:Task2vs1	0.24	0.18	1.34	.180
GroupAH-regularvsSA:Task3vs1	−0.40	0.21	−1.94	.053
GroupAH-intensivevsSA:Task3vs1	0.19	0.22	0.88	.381
TimeT2vsT1:GroupAH-regularvsSA:Task2vs1	0.18	0.18	0.98	.326
TimeT2vsT1:GroupAH-intensivevsSA:Task2vs1	0.07	0.19	0.34	.736
TimeT2vsT1:GroupAH-regularvsSA:Task3vs1	0.33	0.21	1.56	.118
TimeT2vsT1:GroupAH-intensivevsSA:Task3vs1	0.24	0.23	1.04	.300

Notes: Model specification:
glmer(Accuracy ~ Time * Group * Task + (1 + Time + Group|Item_number) + (1 + Time + Task|SubjectNo)).
Tasks 1, 2, and 3 refer to the lexical access, grammatical processing, and semantic processing task, respectively; T1 and T2 refer to the pretest and the posttest, respectively.

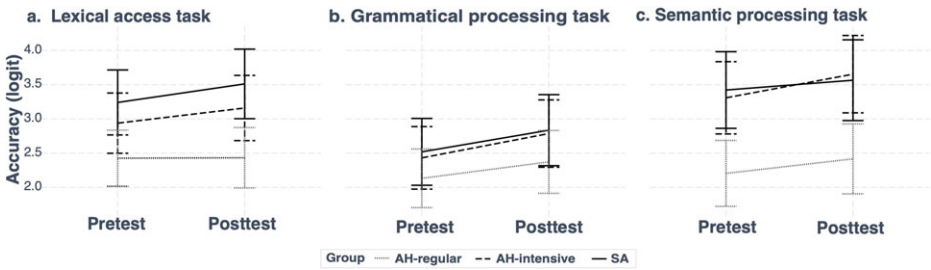

FIGURE 2. Interaction between Time, Group, and Task effects in the processing accuracy model with performance on different linguistic tasks in different panels. Y-axis does not start from zero. Error bars represent standard errors.

between the AH-regular and SA groups in the lexical access task was larger than that in the grammatical processing task (β = 0.44, *SE* = 0.17, z = 2.55, p = .011), but was slightly smaller than that in the semantic processing task (β = −0.40, *SE* = 0.21, z = −1.94, p = .053).

As for the accuracy difference between pretest and posttest, participants had higher accuracy in the posttest than in the pretest across tasks (β = 0.23, *SE* = 0.04, z = 5.15, p < .001). The general accuracy improvement did not differ across either task (β = 0.15,

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

$SE = 0.10$, $z = 1.46$, $p = .146$; $\beta = 0.09$, $SE = 0.11$, $z = 0.80$, $p = .424$) or group ($\beta = -0.09$, $SE = 0.08$, $z = -1.11$, $p = .269$; $\beta = 0.05$, $SE = 0.09$, $z = 0.58$, $p = .565$). None of the three-way interactions was significant.

PROCESSING SPEED

As Table 5 and Figure 3 show, at the *pretest*, similar to the accuracy model, the SA group outperformed the AH-regular group ($\beta = 0.10$, $SE = 0.02$, $t = 4.30$, $p < .001$), but did not differ significantly from the AH-intensive group ($\beta = 0.00$, $SE = 0.03$, $t = 0.16$, $p = .876$) in terms of processing speed. RT performance at the pretest differed across tasks ($\beta = 0.29$, $SE = 0.08$, $t = 3.58$, $p < .001$; $\beta = 0.21$, $SE = 0.09$, $t = 2.39$, $p = .018$), which was expected as the stimuli duration differs across the tasks. Furthermore, the difference between the AH-regular and SA groups in the lexical access task was also larger than that in the grammatical processing task ($\beta = -0.08$, $SE = 0.03$, $t = -3.15$, $p = .002$) and did not differ significantly from that of the semantic processing task ($\beta = 0.04$, $SE = 0.03$, $t = 1.36$, $p = .174$). Similarly, the difference between the AH-intensive and SA group in the lexical access task was larger than that in the grammatical processing task ($\beta = -0.06$, $SE = 0.03$, $t = -2.17$, $p = .031$), but did not differ significantly from that of the semantic processing task ($\beta = -0.06$, $SE = 0.03$, $t = -1.95$, $p = .053$).

TABLE 5. Fixed-effect estimates of RT performance of participant groups in the three processing tasks

	B	*SE*	*t*	*p*
(Intercept)	3.20	0.45	7.07	< .001
TimeT2vsT1	−0.12	0.00	−49.89	< .001
GroupAH-regularvsSA	0.10	0.02	4.30	< .001
GroupAH-intensivevsSA	0.00	0.03	0.16	.876
Task2vs1	0.29	0.08	3.58	< .001
Task3vs1	0.21	0.09	2.39	.018
log(audio_duration)	0.64	0.06	10.03	< .001
Trial_number	−0.00	0.00	−11.03	< .001
TimeT2vsT1:GroupAH-regularvsSA	−0.03	0.01	−4.55	< .001
TimeT2vsT1:GroupAH-intensivevsSA	0.03	0.01	4.98	< .001
TimeT2vsT1:Task2vs1	0.01	0.01	1.70	.089
TimeT2vsT1:Task3vs1	0.01	0.01	1.48	.139
GroupAH-regularvsSA:Task2vs1	−0.08	0.03	−3.15	.002
GroupAH-intensivevsSA:Task2vs1	−0.06	0.03	−2.17	.031
GroupAH-regularvsSA:Task3vs1	0.04	0.03	1.36	.174
GroupAH-intensivevsSA:Task3vs1	−0.06	0.03	−1.95	.053
TimeT2vsT1:GroupAH-regularvsSA:Task2vs1	0.09	0.01	6.55	< .001
TimeT2vsT1:GroupAH-intensivevsSA:Task2vs1	0.02	0.01	1.28	.201
TimeT2vsT1:GroupAH-regularvsSA:Task3vs1	0.04	0.01	2.94	.003
TimeT2vsT1:GroupAH-intensivevsSA:Task3vs1	0.01	0.01	0.45	.655

Note: Model specification:
lmer(log_RT ~ Time * Group * Task + log_audio_duration + Trial_number + (1 + Task|SubjectNo) + (1+Group|Item_number)). Tasks 1, 2, and 3 refer to the lexical access, grammatical processing, and semantic processing task, respectively; T1 and T2 refer to the pretest and the posttest, respectively.

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

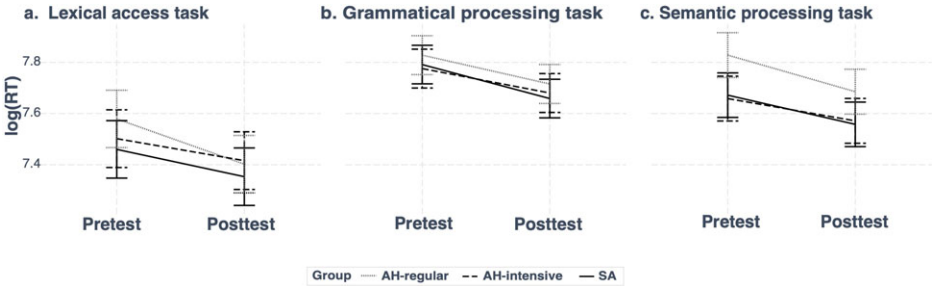

FIGURE 3. Interaction between Time, Group, and Task effects in the processing speed model with performance on different linguistic tasks in different panels. Y-axis does not start from zero. Error bars represent standard errors.

As for the RT difference between pretest and posttest, participants generally responded faster in the posttest than in the pretest ($\beta = -0.12$, $SE = 0.00$, $t = -49.89$, $p < .001$). The progress that participants made in RT did not differ significantly across tasks ($\beta = 0.01$, $SE = 0.01$, $t = 1.70$, $p = .089$; $\beta = 0.01$, $SE = 0.01$, $t = 1.48$, $p = .139$). The AH-regular group made more progress in processing speed than the SA group ($\beta = -0.03$, $SE = 0.01$, $t = -4.55$, $p < .001$), who in turn made more progress than the AH-intensive group ($\beta = 0.03$, $SE = 0.01$, $t = 4.98$, $p < .001$). The degree of progress made by the groups in RT was affected by tasks. More specifically, the difference between the SA and the AH-regular group, in terms of progress in RT performance, in the lexical access task was significantly different from that in the grammatical processing ($\beta = 0.09$, $SE = 0.01$, $t = 6.55$, $p < .001$) and semantic processing tasks ($\beta = 0.04$, $SE = 0.01$, $t = 2.94$, $p < .01$). To clarify the three-way interactions between Time, Group, and Task effects, we split the dataset by task and then fitted models for each task dataset (see Appendix Table A2). The AH-regular group made more progress than the SA group in the lexical access task ($\beta = -0.07$, $SE = 0.01$, $t = -7.02$, $p < .001$), which was also observed in the semantic processing task but with a smaller effect size ($\beta = -0.03$, $SE = 0.01$, $t = -3.04$, $p = .002$). However, the reverse pattern for the same group comparison was found in the grammatical processing task ($\beta = 0.02$, $SE = 0.01$, $t = 1.97$, $p = .049$).

PROCESSING STABILITY

As is shown in Table 6 and Figure 4, the three participant groups did not differ significantly in processing stability in the *pretest* ($\beta = 0.01$, $SE = 0.01$, $t = 1.56$, $p = .121$; $\beta = 0.01$, $SE = 0.01$, $t = 1.29$, $p = .198$). RTs in the lexical access task were more stable than those in the grammatical processing task ($\beta = 0.04$, $SE = 0.01$, $t = 6.40$, $p < .001$), and less stable than those in the semantic processing task ($\beta = -0.07$, $SE = 0.01$, $t = -10.78$, $p < .001$). Moreover, the difference between the AH-regular and SA groups on the lexical access task was significantly different from that on the grammatical processing task ($\beta = -0.03$, $SE = 0.02$, $t = -2.09$, $p = .037$), but did not differ significantly from that of the semantic processing task ($\beta = 0.00$, $SE = 0.02$, $t = 0.27$, $p = .789$).

As for the CV difference between pretest and posttest, participants' RTs were generally more stable in the posttest than in the pretest ($\beta = -0.02$, $SE = 0.00$, $t = -5.30$, $p < .001$).

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

TABLE 6. Fixed-effect estimates of CV performance of the participant groups in the three processing tasks

	B	*SE*	*t*	*p*
(Intercept)	0.33	0.00	91.41	< .001
TimeT2vsT1	−0.02	0.00	−5.30	< .001
GroupAH-regularvsSA	0.01	0.01	1.56	.121
GroupAH-intensivevsSA	0.01	0.01	1.29	.198
Task2vs1	0.04	0.01	6.40	< .001
Task3vs1	−0.07	0.01	−10.78	< .001
TimeT2vsT1:GroupAH-regularvsSA	−0.00	0.01	−0.18	.854
TimeT2vsT1:GroupAH-intensivevsSA	0.01	0.01	1.00	.317
TimeT2vsT1:Task2vs1	0.04	0.01	4.02	< .001
TimeT2vsT1:Task3vs1	0.01	0.01	0.64	.526
GroupAH-regularvsSA:Task2vs1	−0.03	0.02	−2.09	.037
GroupAH-intensivevsSA:Task2vs1	0.00	0.02	0.27	.790
GroupAH-regularvsSA:Task3vs1	0.00	0.02	0.27	.789
GroupAH-intensivevsSA:Task3vs1	0.02	0.02	1.01	.313
TimeT2vsT1:GroupAH-regularvsSA:Task2vs1	0.00	0.02	0.21	.835
TimeT2vsT1:GroupAH-intensivevsSA:Task2vs1	−0.04	0.02	−1.85	.065
TimeT2vsT1:GroupAH-regularvsSA:Task3vs1	0.00	0.02	0.15	.882
TimeT2vsT1:GroupAH-intensivevsSA:Task3vs1	−0.02	0.02	−1.04	.298

Note: Model specification:
lmer(CV_pp ~ Time * Group * Task +(1 + Time|SubjectNo)).Tasks 1, 2, and 3 refer to the lexical access, grammatical processing, and semantic processing task, respectively; T1 and T2 refer to the pretest and the posttest, respectively.

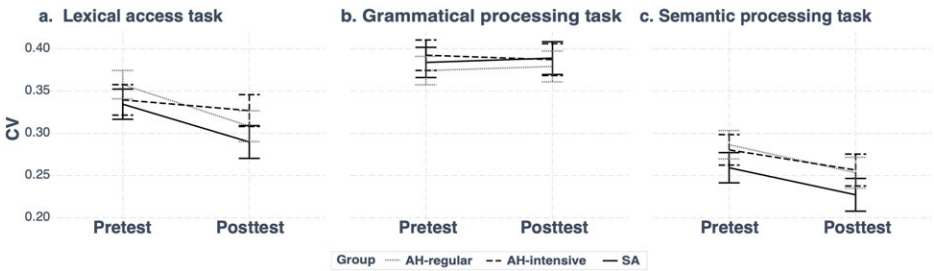

FIGURE 4. Interaction between Time, Group, and Task effects in the processing stability model with performance on different linguistic tasks in different panels. Y-axis does not start from zero. Error bars represent standard errors.

The progress participants made in the lexical access task did not differ significantly from that in the semantic processing task ($\beta = 0.01$, $SE = 0.01$, $t = 0.64$, $p = .526$), but was larger than that in the grammatical processing ($\beta = 0.04$, $SE = 0.01$, $t = 4.02$, $p < .001$). The stability difference between the pretest and posttest was not modulated by group ($\beta = -0.00$, $SE = 0.01$, $t = -0.18$, $p = .854$; $\beta = 0.01$, $SE = 0.01$, $t = 1.00$, $p = .317$). This model had no significant three-way interaction between Time, Group, and Task effects.

In summary, with regard to baseline proficiency, the AH-intensive group had the largest vocabulary size at the pretest, followed by the SA group who in turn outperformed

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

the AH-regular group. Moreover, the AH-intensive and SA groups did not differ from each other but outperformed the AH-regular group in terms of processing efficiency at the pretest. As for the performance difference between the pretest and the posttest, the SA and AH-regular groups made comparable progress in vocabulary size, but the AH-intensive group did not make any significant progress. Meanwhile, the AH-regular group made more progress in processing speed than the SA groups, who made more progress than the AH-intensive group. However, the progress in accuracy and stability of processing was not significantly different among these three learner groups.

DISCUSSION

THE EFFECT OF LEARNING CONTEXTS ON VOCABULARY SIZE

Our first research question was whether the improvement in auditory vocabulary over the course of an academic year is conditioned by learning context. We found that the SA and AH-regular groups made comparable progress in vocabulary size, and their progress was larger than that of the AH-intensive group. Post-hoc analyses on vocabulary improvement in each of the separate groups showed that the AH-intensive group did not make significant progress but both the SA and AH-regular groups did. These results can be broken down into two SA-AH comparisons. First, although the SA group made significantly more progress than the AH-intensive group, the significance may be driven either by the fact that the AH-intensive group did not improve, or by the advantageous effects of the SA context on vocabulary acquisition as compared to the AH context. Note that even at posttest, the mean vocabulary scores of the AH-intensive group were still numerically higher than those of the SA group. Thus, it cannot be concluded that the SA context was more effective in facilitating vocabulary acquisition, based on the comparison between the SA and the AH-intensive groups. Second, vocabulary improvement of the SA group over an academic year abroad did not differ significantly from that of the AH-regular group. However, it is unclear whether this similar improvement pattern for these two groups should be attributed (partly) to differences in baseline vocabulary size.

Taken together, these results do not provide clear evidence to support our initial hypothesis about the facilitative role of the SA context in vocabulary acquisition relative to the AH context. Therefore, even though adult L2 learners may benefit more from the extra incidental learning opportunities provided by SA context relative to the AH context, the magnitude of the learning-context advantage on vocabulary improvement seems small or nonexistent. One possible reason is that SA learners may have faced problems with social integration, leading to a low degree of immersion (see Coleman, 2013 for a model of socialization while abroad). That is, the supposedly rich input, output, and interaction in the SA context may turn out to be shallow due to integration problems, which may make learning advanced vocabulary difficult. However, it is conceivable that the SA group may have acquired more vocabulary used in their specific discipline (e.g., architecture, chemistry of philosophy) during class and more vocabulary relevant to their immediate living experience (e.g., names of grocery items or cooking utensils) out of class than their AH peers, but such vocabulary gains may not be effectively detected by the PPVT™-4.

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

The present study complements studies on the impact of learning contexts on vocabulary acquisition in reading, writing, and speaking activities (e.g., Briggs, 2015; DeKeyser, 1991; Dewey, 2004, 2008; Ife et al., 2000; Llanes & Muñoz, 2009; Milton & Meara, 1995). These studies also present a mixed picture of SA effects, especially on reading vocabulary development. More specifically, some studies reported substantially greater gains in reading vocabulary in SA contexts compared to AH contexts (e.g., Milton & Meara, 1995), whereas others found no significant difference between intensive domestic instruction and SA contexts (e.g., Dewey, 2004, 2008; Serrano et al., 2011). As for productive vocabulary used in speaking and writing, no significant advantage of SA over AH contexts was found concerning the acquisition of new words (Collentine, 2004; Freed et al., 2003).

In contrast to Dewey (2008) and Serrano et al. (2011), who reported comparable vocabulary development in intensive domestic immersion and SA contexts, the improvement patterns of the AH-intensive and SA groups in the present study seem to suggest that the intensive domestic program may not be effective in enhancing listening vocabulary for relatively advanced learners. In other words, the fact that there was no significant difference in vocabulary size of the AH-intensive group between pretest and posttest could mean that this group had reached a plateau in auditory vocabulary acquisition. However, because the average vocabulary score of the AH-intensive group is far from the ceiling performance according to the PPVT™-4 manual, there should be plenty of room for vocabulary growth. Therefore, the lack of improvement of auditory vocabulary in the AH context is unexpected. Previous studies (e.g., Han, 2013; Han & Odlin, 2006; Selinker, 1972) have speculated on a stabilization or fossilization phenomenon where L2 language proficiency stops improving regardless of abundant target-language exposure. However, these studies have mostly been carried out in study/residence abroad contexts in which learners are immersed in their L2. We speculate that learners could also reach a point of stabilization in auditory vocabulary acquisition in AH contexts with generally impoverished target language exposure, but note that measurement at two time points does not allow a firm conclusion about stabilization.

THE EFFECT OF LEARNING CONTEXTS ON PROCESSING EFFICIENCY

Our second research question addressed the question of whether the SA and the AH learning contexts differed in facilitating the development of L2 listening processing skills. Participant groups all improved in terms of accuracy and stability of processing, but there were no significant group differences in the amount of improvement for these two measures. As for the speed of processing, the SA group made more progress than the AH-intensive group and less improvement than the AH-regular group. We interpret these results as follows. Firstly, the comparison between the SA group and the AH-intensive group suggests that the SA learners improved their speed of processing more rapidly than their AH peers. As processing efficiency of these two groups did not differ at the pretest, it is likely the observed effects are not due to baseline confounds, but rather should be explained by the effect of learning context. Secondly, the fact that the SA group showed less improvement than the AH-regular group in the speed of processing seems to contradict our hypothesis that the SA context would facilitate processing efficiency better than the AH context. However, steep learning curves for low-proficiency participants in

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

reaction time were commonly observed in previous studies (e.g., van den Bosch et al., 2019). Because the AH-regular group had lower proficiency than the AH-intensive and the SA group at pretest, it can be argued that the AH-regular group improved relatively fast due to their lower starting level at the beginning of the academic year. Therefore, the effect of learning context, substantiated by the comparison between SA and AH-regular groups, should be considered in light of baseline proficiency.

Therefore, provided equal starting levels, studying abroad is more beneficial for enhancing L2 processing speed over remaining at home. This agrees with our initial hypothesis about the facilitative effect of SA learning contexts on processing efficiency. The findings of the present study complement, though not always align with, those of previous studies on fluency (e.g., Freed et al., 2004; Sasaki, 2007; Segalowitz & Freed, 2004), another aspect of language processing. Previous studies have usually associated greater gains in fluency with SA contexts relative to AH contexts. Freed et al. (2004) compared oral fluency development (L1 English, L2 French) in AH, SA, and domestic immersion contexts over one semester, and found that the SA group improved more than the AH group but less than the domestic immersion group. Note, however, that the questionnaires of that study revealed that the domestic immersion group surprisingly used the target L2 French more than the SA group. In contrast, Sasaki (2007) compared changes in the writing of SA learners and domestic English majors (L1 Japanese, L2 English) during 1 year and found that the SA group improved their English writing fluency but the AH group did not. The present study provides evidence that similar to fluency development in oral and written production, the speed of L2 processing in listening comprehension develops more rapidly in an SA context than an AH context. This suggests that the development of L2 processing skills is subject to the effect of learning contexts.

Furthermore, the advantage of the SA group over the AH-intensive group in facilitating the development of L2 processing efficiency was not constrained to specific linguistic processes. This means that the effect of learning contexts on improving L2 processing efficiency manifests in all three linguistic processes studied (i.e., lexical access, grammatical processing, and semantic interpretation). Previous studies on the effect of learning context on either lexical access or grammatical processing, however, seem to present a different picture. Segalowitz and Freed (2004) tested SA and AH learners with a semantic classification task in a pretest and a posttest and found no differential gains in lexical access as a function of learning contexts. Similarly, Isabelli-García (2010) and Pliatsikas and Marinis (2013a) also reported no difference between these two contexts in grammar acquisition, such as the acquisition of gender agreement and past tense. The present study contradicts with these studies in that we found learning-context effects on language processing at lexical, morphosyntactic, and semantic levels. However, the previously mentioned studies all measured L2 processing in reading, while the present study targeted that in listening. The inconsistent findings between the previous studies and the present study suggest that the SA context may have differential effects on L2 processing in listening and reading activities. This may be related to the fact that formal instruction in AH contexts usually focuses on reading rather than listening. Thus, when learners move to an SA context, their L2 listening ability is likely to develop more rapidly than their reading. In addition, comparing task performance at the pretest and the posttest, we found the grammatical processing task to show the least difference between the participant

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

groups. This aligns with and reconfirms Yu et al. (2020) argument that, compared to lexical access and semantic processing, grammar sensitivity is the area that least distinguished the proficiency of nonnative groups. However, though between-group differences in L2 processing were enlarged for specific linguistic tasks at both testing times, performance on the different linguistic tasks developed similarly over time. One possible reason is that the improvement in L2 processing over one academic year may not be big enough to show the finer differences in the development of linguistics processes, or that the measurement used in the present study may not be sensitive enough to show such differences.

IMPLICATIONS, LIMITATIONS, AND FUTURE DIRECTIONS

This study has pedagogical implications for language learning in both SA and AH contexts.

1. Compared to the domestic non-English-major postgraduates and SA learners, the domestic English-major postgraduates made the least progress in both vocabulary knowledge and processing efficiency. Therefore, the AH context seems to be less effective for L2 listening development of relatively high-proficiency learners. The curriculum of domestic English majors, especially for postgraduates, may need to incorporate more opportunities for interactive language practice. Otherwise, a sojourn abroad might be conducive to further progress.
2. The fact that the SA group did not differ from the AH-regular group in vocabulary improvement supports the view that the role of incidental vocabulary learning may be limited. SA learners ought to seek systematic vocabulary learning activities to achieve greater learning outcomes.
3. We recommend including language processing tests in L2 learning and teaching practice. Such tests offer an accurate measurement of L2 processing efficiency that reveals learners' L2 strengths and weaknesses from a skill acquisition perspective, thus providing learners, teachers, and researchers more insight on SLA.

This study has certain limitations. Firstly, vocabulary knowledge of participant groups differed from each other at the pretest, which may confound the effect of learning contexts to some degree. As it is difficult to manipulate baseline proficiency levels of students in natural learning contexts, previous studies often tend to ignore the effect of baseline proficiency level and directly compare SA and AH learners. Our study addresses this problem by using proper statistical analysis methods (i.e., LMMs) and taking baseline proficiency differences into account when interpreting our results. Note that these analytical methods are a statistical control of baseline proficiency differences, but do not account for other potential differences between the groups. To investigate whether our vocabulary developmental pattern results should indeed be attributed to baseline proficiency differences, replication with participant groups matched on baseline vocabulary knowledge is recommended. Secondly, it should be noted that performance on the processing tasks is sensitive to repetition such that the general progress participants made may be influenced by test-retest effects. However, it is the relative between-group difference in progress that is the major concern of this study, instead of the absolute amount of progress.

To our knowledge, this is the first study to examine the effect of SA and AH learning contexts on the development of L2 listening proficiency from both language knowledge and language processing perspective. Future studies are encouraged to further this

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

investigation in other learning contexts, such as heritage language learning, residence abroad, domestic immersion, and computer-assisted learning contexts, to shed more light on the relation between L2 listening development and learning contexts. Moreover, the linguistic benefits of the SA context may be affected by the duration of an SA experience. The effect of SA duration, especially the contrast between long- and short-term study abroad, should be addressed by future research.

CONCLUSIONS

We found that, in terms of vocabulary gains, the SA group made more progress than the AH-intensive group, but did not differ significantly from the AH-regular group. However, it is difficult to tell whether this differential vocabulary growth should be attributed to the learning context, or may be due to different starting levels. At the same time, the SA group made more progress than the AH-intensive group but less progress than the AH-regular group in speed of lexical, morphosyntactic, and semantic processing. Because the SA and AH-intensive groups started off with equal processing efficiency levels at pretest while the SA and AH-regular groups did not, we argue that the difference between the SA and AH-intensive groups in terms of processing efficiency improvement should be attributed to the effect of learning context. More specifically, this suggests that the SA context facilitates the acquisition of processing skills (processing speed in particular) better than the AH context. To sum up, this study demonstrates that study abroad is an effective intervention for developing L2 processing efficiency with less clear effects on vocabulary acquisition.

REFERENCES

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press.

Anderson, J. R. (2015). *Cognitive psychology and its implications*. Worth Publishers.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Brecht, R. D., & Robinson, J. L. (1995). On the value of formal instruction in study abroad: Student reactions in context. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 317–334). John Benjamins.

Briggs, J. G. (2015). Out-of-class language contact and vocabulary gain in a study abroad context. *System*, *53*, 129–140.

Coleman, J. A. (2013). Researching whole people and whole lives. In C. Kinginger (Ed.), *Social and cultural aspects of language learning in study abroad* (pp. 17–44). John Benjamins.

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

Cubillos, J. H., Chieffo, L., & Fan, C. (2008). The impact of short-term study abroad programs on L2 listening comprehension skills. *Foreign Language Annals*, *41*, 157–186.

Cutler, A., & Charles, C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford University Press.

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

DeKeyser, R. M. (1991). The semester overseas: What difference does it make? *ADFL Bulletin*, *22*, 42–48.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, *19*, 195–221.

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

DeKeyser, R. M. (Ed.). (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press.

DeKeyser, R. M. (2015). Skill acquisition theory. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 94–112). Routledge.

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.

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

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*. Pearson Assessments.

Dwyer, M. (2004). More is better: The impact of study abroad program duration. *Frontiers: The Interdisciplinary Journal of Study Abroad*, *10*, 151–163.

Ellis, R., Tanaka, Y., & Yamazaki, A. (1994). Classroom interaction, comprehension, and the acquisition of L2 word meanings. *Language Learning*, *44*, 449–491.

Foster, P., Bolibaugh, C., & Kotula, A. (2014). Knowledge of nativelike selections in a L2: The influence of exposure, memory, age of onset, and motivation in foreign language and immersion settings. *Studies in Second Language Acquisition*, *36*, 101–132.

Freed, B., So, S., & Lazar, N. A. (2003). Language learning abroad: How do gains in written fluency compare with gains in oral fluency in French as a second language? *ADFL Bulletin*, *34*, 34–40.

Freed, B. F. (Ed.). (1995). *Second language acquisition in a study abroad context*. John Benjamins.

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.

Goss, B. (1982). Listening as information processing. *Communication Quarterly*, *30*, 304–307.

Håkansson, G., & Norrby, C. (2010). Environmental influence on language acquisition: Comparing second and foreign language acquisition of Swedish. *Language Learning*, *60*, 628–650.

Han, Z. (2013). Forty years later: Updating the fossilization hypothesis. *Language Teaching*, *46*, 133–171.

Han, Z., & Odlin, T. (Eds.). (2006). *Studies of fossilization in second language acquisition*. Multilingual Matters.

Horst, M. (2005). Learning L2 vocabulary through extensive reading: A measurement study. *The Canadian Modern Language Review*, *61*, 355–382.

Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223.

Howard, M. (2001). The effects of study abroad on the L2 learner's structural skills: Evidence from advanced learners of French. *EUROSLA Yearbook*, *1*, 123–141.

Hulstijn, J. H. (2003). Incidental and intentional learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Blackwell.

Hulstijn, J. H., Van Gelderen, A., & Schoonen, R. (2009). Automatization in second language acquisition: What does the coefficient of variation tell us? *Applied Psycholinguistics*, *30*, 555–582.

Ife, A., Vives Boix, G., & Meara, P. (2000). The impact of study abroad on vocabulary development among different proficiency groups. *Spanish Applied Linguistics*, *4*, 55–84.

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

Jacobs, A., Fricke, M., & Kroll, J. F. (2016). Cross-language activation begins during speech planning and extends into second language speech. *Language learning*, *66*, 324–353.

Jiang, N. (2018). *Second language processing: An introduction*. Routledge.

Kersten, K. (2010). *ELIAS: Early language and intercultural acquisition studies. Final report: Public part*. ELIAS.

Knoch, U., Rouhshad, A., Oon, S. P., & Storch, N. (2015). What happens to ESL students' writing after three years of study at an English medium university? *Journal of Second Language Writing*, *28*, 39–52.

Kroll, J. F., Dussias, P. E., & Bajo, M. T. (2018). Language use across international contexts: Shaping the minds of L2 speakers. *Annual Review of Applied Linguistics*, *38*, 60–79.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 1–26.

Lafford, B. A. (2004). The effect of the context of learning on the use of communication strategies by learners of Spanish as a second language. *Studies in Second Language Acquisition*, *26*, 201–225.

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

Leong, C.-H. (2007). Predictive validity of the multicultural personality questionnaire: A longitudinal study on the socio-psychological adaptation of Asian undergraduates who took part in a study-abroad program. *International Journal of Intercultural Relations*, *31*, 545–559.

Lim, H., & Godfroid, A. (2015). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, *36*, 1247–1282.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*, 1507–1515.

Llanes, À., & Muñoz, C. (2009). A short stay abroad: Does it make a difference? *System*, *37*, 353–365.

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

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Maeder-Qian, J. (2018). Intercultural experiences and cultural identity reconstruction of multilingual Chinese international students in Germany. *Journal of Multilingual and Multicultural Development*, *39*, 576–589.

Marqués-Pascual, L. (2011). Study abroad, previous language experience, and Spanish l2 development. *Foreign Language Annals*, *44*, 565–582.

Martinsen, R. A., Alvord, S. M., & Tanner, J. (2014). Perceived foreign accent: Extended stays abroad, level of instruction, and motivation. *Foreign Language Annals*, *47*, 66–78.

Matsumura, S. (2001). Learning the rules for offering advice: A quantitative approach to second language socialization. *Language Learning*, *51*, 635–679.

Mayer, R. (2009). *Multimedia learning*. Cambridge University Press.

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

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

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research*, *17*, 164–187.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*, 31–55.

Pliatsikas, C., & Marinis, T. (2013a). Processing of regular and irregular past tense morphology in highly proficient second language learners of English: A self-paced reading study. *Applied Psycholinguistics*, *34*, 943–970.

Pliatsikas, C., & Marinis, T. (2013b). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, *16*, 167–182.

Plonsky, L. (2011). The effectiveness of second language strategy instruction: A meta-analysis. *Language Learning*, *61*, 993–1038.

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Sasaki, M. (2007). Effects of study-abroad experiences on EFL writers: A multiple-data analysis. *The Modern Language Journal*, *91*, 602–620.

Sasaki, M. (2011). Effects of varying lengths of study-abroad experiences on Japanese EFL students' L2 writing ability and motivation: A longitudinal study. *TESOL Quarterly*, *45*, 81–105.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363.

Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 382–408). Blackwell.

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

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*, 173–199.

Selinker, L. (1972). Interlanguage. *IRAL—International Review of Applied Linguistics in Language Teaching*, *10*, 209–232.

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.

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

Swanborn, M. S., & De Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research*, *69*, 261–285.

Swanborn, M. S., & De Glopper, K. (2002). Impact of reading purpose on incidental word learning from context. *Language Learning*, *52*, 95–117.

Taatgen, N. A., & Lee, F. J. (2003). Production compilation: A simple mechanism to model complex skill acquisition. *Human Factors*, *45*, 61–76.

Taguchi, N. (2011). The effect of L2 proficiency and study-abroad experience on pragmatic comprehension. *Language Learning*, *61*, 904–939.

van den Bosch, L. J., Segers, E., & Verhoeven, L. (2019). The role of linguistic diversity in the prediction of early reading comprehension: A quantile regression approach. *Scientific Studies of Reading*, *23*, 203–219.

Varela, O. E. (2017). Learning outcomes of study-abroad programs: A meta-analysis. *Academy of Management Learning & Education*, *16*, 531–561.

Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? *Reading in a Foreign Language*, *15*, 130–163.

Waters, G. S., Caplan, D., & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology*, *12*, 1–30.

Weist, R. M. (2002). Space and time in first and second language acquisition: A tribute to Henning Wode. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *The integrated view of language development: Papers in honor of Henning Wode* (pp. 79–109). WVT Wissenschaftlicher Verlag Trier.

Williams, T. R. (2005). Exploring the impact of study abroad on students' intercultural communication skills: Adaptability and sensitivity. *Journal of Studies in International Education*, *9*, 356–371.

Yu, X., Janse, E., & Schoonen, R. (2020). *Breaking down listening comprehension: How does L2 exposure affect vocabulary knowledge and processing efficiency in EFL learners*. Manuscript submitted for publication.

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534

APPENDIX

TABLE A1. Fixed-effect estimates of vocab models (split by groups)

	β	*SE*	*t*	*p*
AH-regular group				
(Intercept)	116.45	3.77	30.91	<.001
TimeT2vsT1	4.87	2.29	2.13	.038
AH-intensive group				
(Intercept)	147.54	2.45	60.31	<.001
TimeT2vsT1	1.35	1.49	0.91	.369
SA group				
(Intercept)	131.57	3.27	40.24	<.001
TimeT2vsT1	7.45	1.80	4.14	<.001

Note: Model specification in lmer(Vocab ~ Time + (1|SubjectNo)).

TABLE A2. Fixed-effect estimates of RT models (split by tasks)

	β	*SE*	*t*	*p*
Lexical access task				
(Intercept)	5.91	1.12	5.27	<.001
TimeT2vsT1	−0.12	0.00	−29.97	<.001
GroupAH-regularvsSA	0.12	0.03	3.60	<.001
GroupAH-intensivevsSA	0.04	0.03	1.21	.228
log_audio_duration	0.18	0.18	1.00	.321
Trial_number	0.00	0.00	0.92	.359
TimeT2vsT1:GroupAH-regularvsSA	−0.07	0.01	−7.02	<.001
TimeT2vsT1:GroupAH-intensivevsSA	0.02	0.01	1.99	.047
Grammatical processing task				
(Intercept)	2.56	0.61	4.19	<.001
TimeT2vsT1	−0.11	0.00	−27.25	<.001
GroupAH-regularvsSA	0.04	0.02	1.50	.136
GroupAH-intensivevsSA	−0.02	0.02	−0.69	.492
log_audio_duration	0.75	0.08	9.09	<.001
Trial_number	−0.00	0.00	−14.13	<.001
TimeT2vsT1:GroupAH-regularvsSA	0.02	0.01	1.97	.049
TimeT2vsT1:GroupAH-intensivevsSA	0.04	0.01	3.78	<.001
Semantic processing task				
(Intercept)	4.29	0.87	4.93	<.001
TimeT2vsT1	−0.11	0.00	−31.00	<.001
GroupAH-regularvsSA	0.16	0.03	6.01	<.001
GroupAH-intensivevsSA	−0.01	0.03	−0.48	.629
log_audio_duration	0.50	0.12	4.33	<.001
Trial_number	−0.00	0.00	−7.08	<.001
TimeT2vsT1:GroupAH-regularvsSA	−0.03	0.01	−3.04	.002
TimeT2vsT1:GroupAH-intensivevsSA	0.03	0.01	3.08	.002

Note: Model specification in lmer(log_RT ~ Time * Group + log_audio_duration + Trial_number + (1|SubjectNo) + (1+Group|Item_number).

Downloaded from https://www.cambridge.org/core. Max-Planck-Institute for Psycholinguisti, on 22 Jul 2021 at 07:28:32, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263120000534