LANGUAGE LEARNING
A Journal of Research in Language Studies
Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Is Second Language Attrition Inevitable After Instruction Ends? An Exploratory Longitudinal Study of Advanced Instructed Second Language Users

Nicole Tracy-Ventura [a], Amanda Huensch [b], Jonah Katz,[c] and Rosamond Mitchell [d]

[a]Oklahoma State University [b]University of Pittsburgh [c]University of California Los Angeles [d]University of Southampton

Abstract: Most second language acquisition (SLA) research has documented the processes involved in learning second/foreign languages, with few studies focusing on the

CRediT author statement–**Nicole Tracy-Ventura**: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision; visualization; writing–original draft, and writing–review & editing. **Amanda Huensch**: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; resources; supervision; visualization; writing–original draft, and writing–review & editing. **Jonah Katz**: data curation; formal analysis; visualization; writing–original draft, and writing–review & editing. **Rosamond Mitchell**: conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; resources; supervision, and writing–review & editing.

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

LANGSNAP 2011–013 was funded by the ESRC (award number RES-062-23-2996); LANGSNAP 3.0 was funded by a Language Learning Small Research Grant, a USF World Faculty Mobility Grant, and a USF World New Researcher Grant; LANGSNAP 6.0 was funded by West Virginia University and the University of Southampton. We are grateful to our colleagues Kevin McManus, Laurence Richard, Patricia Romero de Mills, the participants, and our research assistants for their contribution to this research.

Correspondence concerning this article should be addressed to Amanda Huensch, Department of Linguistics, University of Pittsburgh, 2816 Cathedral of Learning, Pittsburgh, PA 15260, USA. E-mail: amanda.huensch@pitt.edu

The handling editor for this manuscript was Charlie Nagle.

DOI: 10.1111/lang.12665

durability of attained second language (L2) skills once instructed learners/users are no longer receiving formal instruction. The current study examines the effects of continued exposure and peak instructional attainment on the long-term evolution of advanced, instructed L2 learners' skills following a longitudinal mixed-methods research design. Participants ($n = 28$) completed an oral proficiency test, an oral interview, and a vocabulary knowledge test at multiple times over an 8-year period, 6 years of which were postinstruction. Results showed that continued exposure contributes to long-term retention (and some further development) of oral proficiency and fluency and that peak attainment at the end of formal instruction is also an important variable for some areas of L2 performance. Additionally, even the participants with limited exposure demonstrated little attrition over time.

Keywords L2 attrition; long-term retention; longitudinal research; learner corpora

Introduction

When university students decide to major in a foreign language, they dedicate years to acquiring advanced proficiency.[1] Some also spend a semester or year studying abroad. Studying another language and living abroad are life-changing experiences with lasting effects on students' personal and professional lives (Kinginger & Zhuang, 2023). Some language graduates, however, end up working in fields requiring little to no use of their second languages (L2s). What happens to those hard-earned language skills when students stop using them regularly? Do they change at all? Are they lost? If they are lost over time, are they lost for good or just temporarily inaccessible? Are some skills impacted more than others? These are important questions for both researchers and language learners. It would also be useful to know what activities help in the long-term maintenance of advanced L2 skills. Unfortunately, at this point, research on this topic is extremely lacking, making it impossible to give students empirically based suggestions.

Unlike first language (L1) acquisition, where everyone with normal cognitive abilities reaches advanced proficiency, L2 acquisition is variable and depends on characteristics such as age, exposure, motivation, and investment. Although much second language acquisition (SLA) research has documented learning processes, far less research has focused on the long-term retention of attained L2 skills, especially after formal instruction has ended (Schmid, 2023); so our understanding of the variables important for long-term retention is limited. For example, little is known about the length of time for which learners can expect to retain their language skills without consistent practice. Bardovi-Harlig and Stringer (2010) posited that attrition may "be thought

of as a normal part of the acquisition process" (p. 39) when learners experience extreme shifts in L2 input and interaction. However, the L2 attrition research to date has produced contradictory results regarding the role of exposure/use (Mickan et al., 2023; Schmid, 2023). Other research has suggested that the level of L2 proficiency attained might be influential, for example, more advanced proficiency will lead to better retention (De Bot & Clyne, 1989; Mehotcheva, 2010; Weltens, 1989). Research on this topic is scarce due to methodological challenges. To investigate the long-term evolution of L2 skills reliably, it is necessary to ensure that L2 skills are measured when learners are most proficient (Bardovi-Harlig & Stringer, 2010) in order to measure any subsequent changes. For many instructed learners, this "peak attainment" would be at the end of formal instruction or after time spent living in a target language context, such as residence/study abroad. To better understand the long-term evolution of L2 proficiency (including attrition, development, and retention) and the variables that play a role, it is necessary to follow participants for several years after language exposure and/or instruction and to gather a variety of linguistic and extralinguistic data. Such observational studies can also help uncover variables that would be important to manipulate in experimental designs.

The current study reports new findings from the longitudinal Languages and Social Networks Abroad Project (LANGSNAP: Mitchell et al., 2017), which initially ran from 2011–2013 to document the learning of L2 French and Spanish among Anglophone undergraduate students ($n = 56$) who were pursuing a 4-year degree in those languages and were required to spend their 3rd year abroad in either France, Mexico, or Spain. In that first phase of the project, participants' linguistic skills were documented at three critical points: before and after a year studying abroad (their 3rd year) and at the end of formal instruction. The data reported in this article are from two follow-up studies, 3 and 6 years after the participants completed their Bachelor's degrees (2016, 2019). All the original participants were invited to participate. A subset volunteered ($n = 33$ in 2016, $n = 28$ in 2019) and completed the same tasks as the original project. Thus, with such data it is possible to explore the long-term evolution of these participants' L2 skills onward from what we refer to as their "peak instructional attainment" (the highest proficiency attained during their studies) and after they experience major shifts in patterns of language use. Here, we focus specifically on three relevant variables (exposure, attainment, and time) in predicting the extent to which participants' L2 proficiency, lexical knowledge, and oral fluency evolve over time.

Literature Review

Attrition of Second Language Skills

Studies focusing on the long-term evolution of bilingual speakers' language skills tend to center on language attrition. The majority of attrition researchers investigate the loss of L1s by people whose L2 becomes their dominant language (e.g., immigrants who move long-term to another country; Schmid, 2011). In studies of L1 attrition, participants who did not move out of the L1 context until aged 13–15+ are typically assumed to have a fully developed L1 (and under these conditions, L1 largely resists attrition; Schmid, 2011). L1 attrition is a thriving area of research, testing important neurolinguistic theories about the nature of language acquisition (Schmid & Köpke, 2017). Such theories can also be tested by investigating L2 retention and attrition; however, research in this area is challenging due to the variable and frequently unstable nature of L2 knowledge compared to L1 knowledge, variables that could mean that the L2 is less resilient when there is a change in learning environment. Because of this variability in L2 end-state proficiency, it is necessary to empirically establish peak attainment before the onset of attrition. However, most studies of L2 attrition use self-report data (Hansen, 1999) or the highest-class level completed (Bahrick, 1984) as peak attainment measures, both of which have been shown to be unreliable methods of assessing language proficiency (Thomas, 2006). Additionally, studies of L2 attrition have primarily used cross-sectional, rather than longitudinal, data. As described by Bardovi-Harlig and Stringer (2010), longitudinal studies of L2 attrition tend to cover only 1–2 years, which is likely not long enough to observe significant changes in proficiency.

Despite their limitations, cross-sectional studies can provide important insights into the timing of L2 attrition and variables that influence it. For example, Bahrick (1984) surveyed a total of 773 participants, the majority of whom ($n = 587$) had stopped learning Spanish between 1 and 50 years prior to the study. Another 146 participants were just completing (or had recently completed) their Spanish classes, and an additional 40 participants had never formally studied Spanish. All participants completed a series of controlled language assessments (reading comprehension, Spanish–English recall vocabulary, Spanish–English recognition vocabulary, English–Spanish recall vocabulary, English–Spanish recognition vocabulary, grammar recall, grammar recognition, idiom recall, idiom recognition, and word order) and a background questionnaire with questions on grades in Spanish classes and use of Spanish and other Romance languages after instruction had ended. Results demonstrated that performance on eight of the 10 variables under investigation

decreased over 3–6 years following the completion of instruction and then stabilized for approximately the next 20 years before decreasing again. The only two variables that did not follow this trend were grammar recall and idiom recognition, both of which continuously declined. These findings suggest that the first 6 years or so after instruction is when evidence of attrition should be apparent for many aspects of proficiency and provide some guidance for a longitudinal study design.

Results of longitudinal studies on L2 attrition in children returnees also suggest that attrition happens quickly after a major change in context. These are children who emigrate temporarily with their families to another country for a period long enough to attend school and have the L2 become their dominant language (e.g., children whose parents come to the United States to pursue a PhD). In this case, the children are considered to have attained native-like proficiency in the L2 (unlike instructed L2 learners). Although longitudinal, such studies are often case studies of individual children. Data collection begins when the children remigrate to their home country and L2 exposure becomes significantly reduced (Flores, 2010, 2015; Tomiyama, 2000). The earliest effects of attrition are typically found around 6 months of reduced exposure (Flores, 2015; Kuhberg, 1992; Tomiyama, 2008). For some participants, it becomes challenging to speak the L2 after 18 months (e.g., Flores, 2015). One of the longest studies (Tomiyama, 2000) demonstrated varying dimensions of attrition 33 months after the single participant returned to Japan, at which point the participant was still able to produce spontaneous speech. Studies of children returnees also demonstrate that variables such as age of onset of input loss and amount of exposure to the L2 postreturn are important predictors of L2 attrition. The younger the child and the less exposure they have to the L2 after returning home, the more they will experience attrition.

In contrast, reduced exposure (input and time) postinstruction has not been shown to be a strong predictor of foreign language (FL) attrition (Mehotcheva, 2010; Murtagh, 2003; Weltens, 1989). Instructed FL learners are quite different from child returnees; they typically do not reach such high levels of proficiency due to the relatively limited amount of input that they receive as well as the different ways in which implicit and explicit learning affect the representation of linguistic knowledge in memory (Schmid & Mehotcheva, 2012). Indeed, Schmid and Mehotcheva have limited the definition of L2 attrition to languages that have been learned naturalistically but that may have declined when learners return to their original L1 setting (including studies of adult learners, such as Hansen, 1999, and Hansen & Chen, 2001). Furthermore, Schmid and Mehotcheva have argued that predictions for attrition may differ in the case of

L2 and FL learning. The participants in the current study are an example of "mixed case" learners who first received explicit FL instruction but who also spent 9 months living abroad in a target language-speaking environment and then returned to their home country and received formal instruction for another year. In this sense, they are quite typical of many students who pursue a degree in languages.

The Role of Exposure and Proficiency Attained in Long-Term Retention of Instructed Second Language Skills

Despite early interest in investigating FL attrition (e.g., Lambert & Freed, 1982; Weltens, 1989), very few studies have examined the variables contributing to long-term retention of instructed L2 skills. In the limited research thus far, reduction in exposure at post instruction period has not been shown to be a strong predictor of L2 attrition (Mehotcheva, 2010; Murtagh, 2003; Schmid, 2023). For example, Mehotcheva (2010) included cross-sectional and longitudinal data collected twice over a 1-year period from five participants to investigate Dutch and German-speaking university students' Spanish FL attrition, 12–72 months after living abroad in Spain. Data came from a variety of measures, including a sociolinguistic interview, a C-test, and a picture-naming test. The main results of the longitudinal data demonstrated that the participants experienced attrition, particularly in accessing lexical items. Results of the cross-sectional data demonstrated that higher initial proficiency (measured via self-assessment) was a significant predictor of language retention. In contrast, attitude and motivation, language (non)contact, or length of exposure postinstruction had little influence. Bahrick (1984) also examined variables that contributed to instructed L2 maintenance and found that the higher the 1) level of Spanish training, 2) mean grade received, and 3) level of training in other Romance languages, the better the retention. Rehearsal/use variables were not significant predictors, although Bahrick's participants reported rarely practicing Spanish.

The conflicting results for instructed L2 learners and children returnees require additional investigation. The neurolinguistic theory of bilingualism (NTB), advanced by Paradis (2004), argues that bilingualism comprises a "system of systems" (p. 130), where active use of part of the system (e.g., a particular language) lowers the neurolinguistic activation threshold for that part and inhibits other competing systems or subsystems. The practical effect is that the most commonly used system(s) are the most accessible, and attrition "is a result of lack of long-term stimulation" (Paradis, 2007, p. 125). In other words, acquired linguistic skills are not permanently lost, but

rather become temporarily inaccessible due to lack of use. Additional research suggests that other individual variables, notably proficiency, might be more important than language use. Based on Bahrick's (1984) analysis, Neisser (1984) claimed that a general critical threshold related to language proficiency may exist. That is, once learners reach a particular proficiency level, their linguistic knowledge becomes less vulnerable to attrition. A small number of researchers in psychology (e.g., Mickan et al., 2020) are conducting experimental studies testing these claims, and preliminary results support the role of interference/competition from another language as an explanatory variable in L2 attrition, that is, causing access/retrieval difficulties. A recent longitudinal study by Mickan et al. (2023) tracked a group of German university students who studied abroad in Spain. Although this study focused only on vocabulary knowledge, they were able to demonstrate that 6 months after returning from Spain, those who had used more Spanish, forgot less. Additionally, those who had more experience over time with Spanish (e.g., studying it longer), and those with a larger vocabulary at the end of study abroad, forgot less.

The Current Study

Schmid and Mehotcheva (2012) argued that longitudinal designs allow for a more reliable analysis of FL attrition because a baseline for participants' language abilities can be documented. As we have seen, Bardovi-Harlig and Stringer (2010) suggested that this baseline should be established at the time of peak attainment. They proposed "a model for the assessment of language attrition and retention" (p. 34) that includes discrete time periods of assessment appropriate for the population under study. For example, in a longitudinal design for learners who study abroad, the first period of this design would be formal language instruction, followed by a period of immersion in the host country, and lastly a postsojourn period in the home country. Bardovi-Harlig and Stringer's model also includes two types of assessments: 1) formal assessments of linguistic knowledge and 2) questionnaires documenting language contact as well as other individual variables such as motivation and attitude.

The current study is based on new data collected from participants in the LANGSNAP Project (Mitchell et al., 2017), initially a 2-year longitudinal project (2011–2013) documenting the learning of L2 French and Spanish among United Kingdom's Anglophone undergraduate students who spent an academic year abroad as part of their Bachelor's degrees and who were expected to achieve advanced proficiency in those languages. As evidence of this, they were required to write an undergraduate dissertation in the target

language. Their oral proficiency was also assessed independently as part of the LANGSNAP study (See the Method section) and demonstrated that upon returning from their year abroad, they had achieved a level representative of advanced proficiency (following the results of Bowden, 2016). Due to the longitudinal nature of the project and the fact that a variety of language skills were analyzed at each data collection wave, data exist documenting the LANGSNAP participants' language abilities after study abroad and at the end of formal instruction, which can be used as a baseline for measuring subsequent changes following their graduation in 2013. In the current study, we refer to these time periods as "peak instructional attainment" to emphasize that the participants were still instructed learners at the time. For most participants, peak instructional attainment was at the end of their 9-month stay abroad. Data collection continued during their final year of university study, and it seemed that for some aspects of their L2, such as those related to utterance fluency (e.g., pausing frequency, see Huensch & Tracy-Ventura, 2017), signs of attrition were appearing, despite participants still receiving formal L2 instruction—a result we attributed to the change in learning context. This finding inspired the current study, which explores the long-term evolution of L2 proficiency and examines how peak instructional attainment and subsequent L2 exposure influence long-term retention. Given our modest sample size, we consider the current study exploratory.

Research Questions

1. To what extent does exposure type (limited vs. consistent) affect retention of oral proficiency, oral fluency, and vocabulary knowledge and use after study abroad and formal instruction?
2. To what extent can peak instructional attainment and exposure type predict long-term retention of L2 skills, and to what extent can exposure type predict long-term retention independently of peak instructional attainment?

Method

Data, analysis code, and all materials can be found at the following link: https://osf.io/sypbn/.

Participants

Participants were the subset of the LANGSNAP participants ($n = 28$) who participated in at least one of the two follow-up data collection waves in 2016 or 2019. As a reminder, the original LANGSNAP study focused on the participants' learning of French and Spanish during their 4-year Bachelor's degree,

which included a mandatory year abroad (year 3; for more details see Mitchell et al., 2017). Three and 6 years after they graduated (2016, 2019), the original ($n = 56$) participants were invited to participate again, with a 50 percent positive response. A majority of the participants were living and working in the United Kingdom, with a handful living abroad (e.g., Australia, Canada, France, Japan). In the first data collection wave (2011), participants' mean age was 20.3 ($SD = 1.4$, range = 19–25). Twenty-five of the 28 participants identified English as their L1 ($n = 1$ Finnish, $n = 2$ Polish), and as typical for languages majors, a majority were female ($n = 25$). For more detailed information about the participants, see chapter 3 of Mitchell et al. (2017). For more information about their experiences postgraduation, see Mitchell et al. (2020).

Materials

Participants completed multiple tasks at each data collection wave, each of which is described next. All materials are available on IRIS (Marsden et al., 2016).

Elicited Imitation Test

L2 proficiency was measured at multiple time points using an Elicited Imitation Test (EIT), originally designed by Ortega et al. (2002) to be used in crosslinguistic research. French (Tracy-Ventura et al., 2014) and Spanish (Bowden, 2016; Ortega, 2000) versions of the test were utilized, both of which have been shown to be valid and reliable measures of general oral proficiency. The EIT comprises 30 test items ranging from 7–19 syllables in length. Both the French and Spanish EITs used in the current study are prerecorded and take approximately 9 minutes to complete. The recording begins with instructions and several practice items in English. Next, participants hear a series of utterances in French/Spanish, followed by a beep, and their task is to repeat as much of the utterance as they can in the time allowed. Participants' responses were audio-recorded.

Oral Interview

A semi-structured oral interview was conducted in the target language at each data collection wave. Depending on the data collection wave—prior to, during, or after residence abroad—a researcher asked questions to gather information about participants' plans for their year abroad, experiences living abroad, lifestyles since graduation, travels, careers, and other topics. Interviews in English were also conducted at the end of the year abroad (2012) and in 2016 and 2019, and these provided additional qualitative data about participants'

Table 1 Languages and social networks abroad project (LANGSNAP) project timeline

Date	Location	Wave	Interview	Proficiency Test (EIT)	Questionnaire
May 2011	Home	Presojourn	L2	Yes	n/a
November 2011	Abroad	In-sojourn 1	L2	No	LEQ
February 2012	Abroad	In-sojourn 2	L2	Yes	LEQ
May 2012	Abroad	In-sojourn 3	L2; English	No	LEQ
October 2012	Home	Postsojourn 1	L2	Yes	LEQ
February 2013	Home	Postsojourn 2	L2	No	LEQ
May 2016	Varied	Postsojourn 3	L2; English	Yes	LEQ; BQ
May 2019	Varied	Postsojourn 4	L2; English	Yes	LEQ; BQ

Note. EIT = Elicited Imitation Test; L2 = second language; LEQ = Language Engagement Questionnaire; BQ = background questionnaire.

lives and aspirations. Interviews covered five or six main questions and lasted approximately 10–20 minutes.

X-Lex

The French and Spanish versions of the X-Lex test (Meara, 2005) were used as a measure of receptive vocabulary. In this task, participants were presented with both real and nonce target language vocabulary items ($n = 120$) and had to indicate whether they recognized the word or not. The real words were taken from a range of frequency bands (up to the 5k band). The final score out of 5,000 was calculated based on positive identification of real words at the different frequency bands plus penalties for recognition of any nonce word.

Questionnaires

Participants completed questionnaires at each data collection wave. During and after residence abroad, participants completed the Language Engagement Questionnaire (LEQ; McManus et al., 2014) covering participants' use of English and any other languages they regularly used. For each language and each item on the LEQ (e.g., *listen to music, write emails, engage in small talk*), participants were asked whether they participated in that practice daily, several times a week, a few times a week, a couple of times a month, rarely, or never. An additional background questionnaire (BQ) was included in 2016 and 2019 to gather supplementary information about participants' life experiences after graduation.

Procedure

Table 1 provides the complete timeline of the LANGSNAP project with the data collection waves and tasks used in the current study highlighted in grey.

Data in the current study come from the collection points at presojourn (before participants went abroad, end of year 2), in-sojourn 3 (at the end of their 9-month stay abroad, end of year 3), postsojourn 1 (5 months after residing abroad, start of year 4), postsojourn 2 (9 months after residing abroad, middle of year 4), postsojourn 3 (4 years after residing abroad and 3 years after graduating from university), and postsojourn 4 (7 years after residing abroad and 6 years after graduating from university).

Data Transcription and Coding

Elicited Imitation Test

The scoring rubrics available on IRIS (including example statements) were used to score the EIT. Each item on the EIT received a score of 0–4 (maximum score 120), with 4 being an exact match and 0 being *unintelligible language* or *single word/no response*. Two members of the research team scored each test and resolved any disagreements. Reliability checks included cross-year comparisons of scoring judgments.

Second Language Interviews

The L2 interviews were transcribed in CLAN (which stands for Computerized Language Analysis) following CHAT conventions (CHAT = Codes for the Human Analysis of Transcripts, MacWhinney, 2000) and checked at least twice by team members. The linguistic variables for lexical diversity and oral fluency were calculated using the MATTR (Moving Average Type-Token Ratio) and FluCALC (Fluency Calculator) commands, respectively, available in CLAN. MATTR is a preferred measure of lexical diversity given that it has been shown to be less sensitive than others to text length (Fergadiotis et al., 2015). The MATTR window length was set at 317, the shortest text length (as recommended by Covington, 2007). FluCALC provides a measure of speech rate in syllables per minute, which was used as the utterance fluency measure in the current analysis. In order to utilize the FluCALC command, it was first necessary to link the audio to the transcript file in CLAN.

Exposure

Using data from the English interview, LEQ, and BQ, participants were coded into two categories (*consistent, limited*) regarding the consistency and intensity of their L2 contact and use postgraduation. For instance, those categorized as *consistent* ($n = 16$) had continued to live abroad (e.g., participants 102 and 173), were using their L2 professionally (e.g., participant 123), or had spent a significant amount of time (i.e., more than 12 months) abroad since graduation

(participants 127 and 161). Those categorized as *limited* ($n = 12$) were participants who did not indicate using the L2 on the LEQ (e.g., participants 111, 122, and 179), those living in the United Kingdom and working in a job that required little to no L2 use (participants 118 and 154), and others.

Analysis

Raw data and analysis code are publicly available at https://osf.io/sypbn/. Because of the limited amount of past research and conflicting results in the existing literature, we consider this research at least partly exploratory. As such, the first step in all of the analyses below was to plot the data and identify patterns. We followed this procedure with multiple regression models examining the effect of time on various linguistic outcomes for participants with more consistent or more limited exposure. In these models, time is difference-coded, so each time variable compares the timepoint in question with the previous one. This approach allows us to track local changes in linguistic variables over time without making any assumptions about the shape of the learning (or "unlearning") curve. Note that this means that time is represented with several variables in the models below, each one quantifying a single "step," rather than being modeled with, for instance, a single continuous variable. In all models below, exposure group is coded as a dummy variable with the limited-exposure group set as the baseline: This allows us to interpret main effects as changes in the limited-exposure group and interactions with the exposure variable as differences between the two groups. These data involve multiple observations of each subject, which is modeled using by-subject random intercepts and slopes in a linear mixed-effect regression, implemented with the lme4 package in R (Bates et al., 2015b). By-subject random intercepts allow each subject to differ in their overall score on the measure at issue and are included in all models. By-subject random slopes allow subjects to differ in how their scores change over time; each such parameter makes model-fitting substantially more complex and requires increasing amounts of data to be accurately estimated. For this reason, we added random slopes one at a time and retained only those that significantly improved model fit according to a chi-square test, following Bates et al. (2015a). This means that the various models presented below have different random-effects structures and, in particular, a different number of random effects associated with the time variables described above. The time steps that are retained in the random-effects structures are essentially those associated with a high degree of between-subject variability. All models are presented below with the random effects that were retained and the full table of fixed effects (those that were systematically manipulated in the experiment).

Fixed effects include an estimated coefficient, standard error for that estimate, and test statistics associated with the Sattherthwaite approximation (*df, t, p*). These last three statistics were implemented using the lmerTest package (Kuznetsova et al., 2017).

Although we have included *p*-values here to help readers assess the robustness of effects, we caution that these should not be taken as binary indicators that an effect is "real" or "not real". Because little prior research exists on the questions examined here and the size of our subject pool is limited, *p*-values make a very limited contribution. It is just as important to focus on the direction of trends in the data and their size.

The final analyses attempt to disentangle some interactions between time, exposure, and peak instructional attainment. For these data, each subject reaches peak instructional attainment only once, and the outcome at the end of the study is only a single measurement, so there is no within-subject effect being studied. To model these data, we used general linear regressions with no random effects and report standard test statistics associated with these simpler models. Peak instructional attainment here is centered around the mean so that other effects can be interpreted with respect to a subject with average peak instructional attainment, rather than zero, which is uninterpretable.

Results

The research questions introduced above concern two main questions: 1) the relationship between type of exposure and retention of oral proficiency, fluency, vocabulary knowledge, and lexical diversity after study abroad and formal instruction ends; and 2) whether any putative effect of exposure type on participants' language skills 6 years after graduation is due to higher peak instructional attainment, or whether there is an effect of continuing exposure on linguistic outcomes independently of peak instructional attainment. We address questions about the effect of exposure on the different linguistic variables in the following subsections. The final section attempts to address whether exposure effects are mediated by peak instructional attainment.

Exposure and Proficiency

Table 2 provides descriptive statistics for proficiency, based on the EIT results, and Figure 1 shows how EIT performance changed over time for those with consistent continuing linguistic exposure and those with more limited continuing exposure. Both groups showed substantial gains in performance between presojourn and postsojourn 1. Subsequently, the two groups diverged to some extent. Those with consistent language exposure showed a slight improvement

Table 2 Descriptive statistics for Elicited Imitation Test performance across groups and timepoints

Condition	Statistic	Presojourn	Postsojourn 1 5 months[a]	Postsojourn 3 4 years	Postsojourn 4 7 years
Limited	*M*	72	97	94	93
	n	12	12	12	12
	SD	13	10	14	13
	95% CI Upper	80	103	102	100
	95% CI Lower	65	91	86	86
Consistent	*M*	76	97	101	104
	n	16	16	13	16
	SD	18	15	9	10
	95% CI Upper	85	104	106	108
	95% CI Lower	67	90	96	99

[a] Time period refers to the amount of time after residence abroad ended.

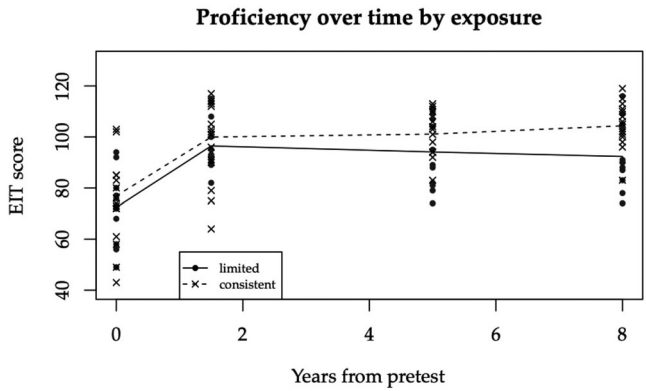

Figure 1 Performance on the Elicited Imitation Task over time, split by group according to amount of continuing L2 exposure. Points are individuals' scores. Lines are locally weighted regressions showing trends in each of the two groups. EIT = Elicited Imitation Test.

over the following 6.5 years, on average, and those with more limited exposure showed a slight decrease in performance, on average.

To examine the time course and robustness of these patterns, we constructed a linear mixed-effects regression model of EIT scores. Fixed effects were exposure type, time, and the interaction between the two. We difference-coded time as a scalar variable, which assessed changes in EIT scores from one time point in our design to the next. Random effects included by-participant intercepts and slopes for two of the three scalar time variables. Results are shown in Table 3.

Table 3 Summary of the Elicited Imitation Test model. Degrees of freedom and *p*-values based on the Satterthwaite approximation. Limited-exposure group set as baseline

Random Effects	Variance	*SD*
Participant (Intercept)	225.2	15.01
Time postsojourn 1 x participant	30.14	5.49
Residual	32.73	5.72

Fixed Effects	Estimate	Std. Error	*df*	*t* value	Pr(>|t|)	95% CI lower	95% CI upper
(Intercept)	72.33	4.64	26.00	15.6	<0.01	63.27	81.4
Exposure consistent	3.79	6.13	26.00	0.62	0.54	–8.2	15.78
Time postsojourn 1 (5 months)	24.58	2.82	40.38	8.71	<0.01	19.13	30.03
Time postsojourn 3 (4 years)	–2.83	2.34	49.50	–1.21	0.23	–7.32	1.65
Time postsojourn 4 (7 years)	–1.25	2.34	49.50	–0.54	0.59	–5.73	3.23
Exposure consistent × Time postsojourn 1	–3.58	3.73	40.38	–0.96	0.34	–10.79	3.63
Exposure consistent × Time postsojourn 3	7.86	3.19	50.34	2.47	0.02	1.74	13.97
Exposure consistent × Time postsojourn 4	2.60	3.19	50.34	0.82	0.42	–3.51	8.72

Before going abroad, the main effect of exposure was small and not robust, indicating that there was no clear or consistent preexisting difference between the two groups that later differed in their continuing exposure. At time postsojourn 1, 5 months after the end of the study abroad program, EIT scores increased substantially (19–30 points on average for the limited-exposure group), and this was significant. The limited-exposure group displayed small decreases in EIT score after this point, but these were not significant. The only significant difference from this timeline in the consistent-exposure group was between the first administration upon return (postsojourn 1) and the follow-up 3.5 years later (postsojourn 3): The limited-exposure group displayed a small (n.s.) decrease in EIT during this span; the consistent-exposure group instead

Table 4 Descriptive statistics for speech rate on the oral interview over time by exposure group

Condition	Statistic	Presojourn	In-sojourn 3	Postsojourn 1 (5 months) [a]	Postsojourn 2 (9 months)	Postsojourn 3 (4 years)	Postsojourn 4 (7 years)
Limited	*M*	136	169	157	165	145	149
	n	12	10	12	12	12	12
	SD	35	31	33	34	32	35
	95% CI Upper	156	188	176	184	163	168
	95% CI Lower	116	150	139	145	126	129
Consistent	*M*	148	176	166	170	175	180
	n	16	16	16	15	13	16
	SD	34	41	40	37	38	46
	95% CI Upper	165	196	186	189	195	203
	95% CI Lower	132	156	147	152	154	158

Note. [a]Time period refers to the amount of time after residence abroad ended.

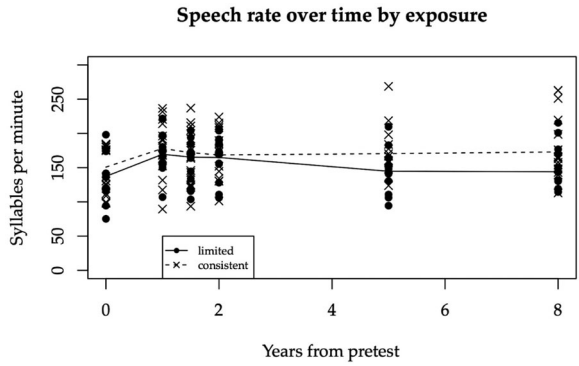

Figure 2 Speech rate over time, split by group according to amount of continuing language exposure. Points are individuals' scores. Lines are locally weighted regressions showing trends in each of the two groups.

improved on the test, which resulted in the significant interaction between exposure and the postsojourn 3 timepoint. The difference between the two groups was estimated at roughly 8 points, but could be as small as 1.7 or as large as 14.

These analyses confirmed that both groups' EIT performance improved postsojourn, that the group with more continuing exposure after return performed better on the EIT than the group with less exposure, and that most of this divergence occurred between 5 months and 4 years postsojourn. As Figure 1 suggests, there was likely even more divergence after this time point, but the effect was smaller and did not reach statistical significance.

Exposure and Fluency

Table 4 provides the descriptive statistics for the fluency variable, operationalized as speech rate in the oral interview. Figure 2 shows how speech rate in syllables per minute changed over time for the two exposure groups, parallel to Figure 1. The six time points were as shown in Table 1. Again, both groups showed substantial gains in performance as expected during their time abroad (between presojourn and in-sojourn 3) and appeared to diverge substantially at the postsojourn 3 time point. Those with consistent language exposure maintained a roughly consistent speech rate over the following 7 years, on average, and those with more limited exposure showed a decrease in speech rate, on average.

The mixed-effects model summarized in Table 5 is similar to the earlier one for EIT but with more timepoints and syllables per minute as the dependent variable. Fixed effects are exposure type, time, and the interaction between the

Table 5 Summary of speech-rate model. Degrees of freedom and *p*-values based on the Sattherthwaite approximation. Limited-exposure group set as baseline

Random Effects	Variance	*SD*
Participant (Intercept)	1,026.75	32.04
Time in-sojourn 3 × participant	722.27	26.88
Time postsojourn 1 × participant	124.26	11.15
Time postsojourn 2 × participant	75.67	8.70
Time postsojourn 3 × participant	284.95	16.88
Residual	162.43	12.75

Fixed effects	Estimate	Std. Error	*df*	*t* value	Pr(>|t|)	95% CI lower	95% CI upper
(Intercept)	136.16	9.95	26.06	13.68	<0.01	116.95	155.61
Exposure consistent	12.23	13.17	26.06	0.93	0.36	–13.19	37.64
Time in-sojourn 3	37.27	9.65	27.82	3.86	<0.01	18.7	55.98
Time postsojourn 1 (5 months)	–16.01	6.59	27.97	–2.43	0.02	–28.72	–3.17
Time postsojourn 2 (9 months)	7.25	5.78	37.75	1.26	0.22	–3.9	18.41
Time postsojourn 3 (4 years)	–19.97	7.13	34.01	–2.8	0.01	–33.79	–6.16
Time postsojourn 4 (7 years)	4.08	5.2	45.83	0.78	0.44	–5.81	14.07
Exposure consistent × Time in-sojourn 3	–9.88	12.6	26.95	–0.78	0.44	–34.26	14.35
Exposure consistent × Time postsojourn 1	6.5	8.45	27.81	0.77	0.45	–9.92	22.76
Exposure consistent × Time postsojourn 2	–5.38	7.7	38.31	–0.7	0.49	–19.99	9.42
Exposure consistent × Time postsojourn 3	28.71	9.68	36.59	2.96	0.01	11.88	46.79
Exposure consistent × Time postsojourn 4	–0.64	7.16	47.23	–0.09	0.93	–14.72	12.13

two. Random effects include by-participant intercepts and slopes for all but one of the scalar time variables.

For the limited-exposure group, speech rate increased substantially by the end of their time abroad (in-sojourn 3; 19–56 syllables/minute). Their speech rate then underwent significant declines by 5 months later (postsojourn 1; 3–29 syllables/minute) and again between 9 months and 4 years later (postsojourn 3; 6–34 syllables/minute). The consistent-exposure group underwent a reasonably similar trajectory, but rather than a decline between times postsojourn 2 and postsojourn 3, they displayed a small increase on average in speech rate, which resulted in the significant interaction between exposure group and time at postsojourn 3.

In summary, similar to EIT performance, both groups' speech rate increased during their time abroad; the group with more continuing exposure had a higher speech rate after return than the group with less exposure, and most of this divergence occurred between 9 months and 4 years after returning.

Exposure and Lexical Diversity

Table 6 displays the descriptive statistics for lexical diversity, operationalized using MATTR, on the oral interview. Figure 3 shows how lexical diversity changed over time across consistent and limited continuing exposure groups. As with the other measures, lexical diversity appeared to increase substantially for both groups during their time abroad (between presojourn and in-sojourn 3). Interestingly, both groups seemed to undergo a slight decline in lexical diversity throughout the first year after their return (between time in-sojourn 3 and postsojourn 2). Lexical diversity scores then showed a slight increase at the next two 3-year intervals, which seemed slightly larger for the consistent exposure group.

Table 7 summarizes a mixed-effects model of lexical diversity parallel to the one of speech rate in the preceding section. Fixed effects are exposure type, time, and the interaction between the two. Random effects include by-participant intercepts and slopes for all but one of the scalar time variables.

Unlike the other two variables examined thus far, none of the time-trends nor the group differences noted in Figure 3 reached statistical significance for lexical diversity scores, except for the rise in diversity at time in-sojourn 3. The MATTR measure is a mean over ratios that range from 0 to 1: Most effects here were smaller than 1% on this scale, and the 95% confidence intervals crossed 0. The two largest effects, estimated between 1 and 2%, were those associated with the initial rise at the end of the speakers' time abroad (in-sojourn 3) and the divergence between groups at timepoint postsojourn 3 roughly 4 years after

Table 6 Descriptive statistics for MATTR (Moving Average Type-Token Ratio) over time by exposure group

Condition	Statistic	Presojourn	In-sojourn 3	Postsojourn 1 (5 months)[a]	Postsojourn 2 (9 months)	Postsojourn 3 (4 years)	Postsojourn 4 (7 years)
Limited	*M*	0.439	0.459	0.455	0.451	0.460	0.462
	n	12	10	12	12	12	12
	SD	0.031	0.032	0.022	0.025	0.027	0.020
	95% CI Upper	0.457	0.479	0.467	0.466	0.475	0.473
	95% CI Lower	0.422	0.439	0.442	0.437	0.445	0.451
Consistent	*M*	0.443	0.465	0.458	0.446	0.470	0.473
	n	16	16	16	15	13	16
	SD	0.024	0.020	0.019	0.029	0.027	0.022
	95% CI Upper	0.455	0.475	0.467	0.461	0.485	0.484
	95% CI Lower	0.431	0.455	0.448	0.431	0.456	0.462

Note. [a]Time period refers to the amount of time after residence abroad ended.

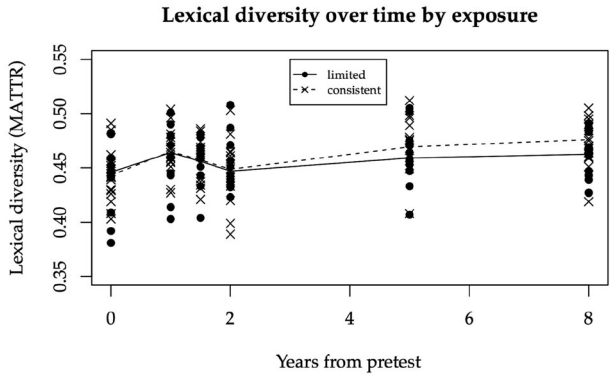

Figure 3 Lexical diversity over time, split by group according to amount of continuing L2 exposure. Points are individuals' scores. Lines are locally weighted regressions showing trends in each of the two groups. MATTR = Moving Average Type-Token Ratio.

the time abroad (interaction between exposure and time postsojourn 3), which is where we found divergence for the other variables as well. Given the noisiness of the data, the interaction at postsojourn 3 did not reach significance at the alpha = 0.05 level. The fact that one of these effects reached the alpha criterion and the other did not is not itself evidence that the two effects are different from one another (Gelman & Stern, 2006); it is just that there is more uncertainty around our estimate of the interaction parameter, as indicated by the confidence intervals. The confidence intervals also indicate that the effect at in-sojourn 3 may be extremely small (as low as 0.001), even if it is robust.

Exposure and Vocabulary Knowledge (X-lex)

Table 8 displays the descriptive statistics for the Exposure and Vocabulary Knowledge test, X-lex, and Figure 4 shows how vocabulary knowledge changed over time across consistent- and limited-exposure groups. As with the other measures, vocabulary increased substantially for both groups during their time abroad (between presojourn and in-sojourn 3). After the return from abroad, scores appeared more or less flat in the limited-exposure group and showed a slight rise in the consistent-exposure group.

The confidence intervals in Table 7 suggest that by the final timepoint in the study, the consistent-exposure group had a substantial vocabulary advantage over the limited-exposure group. But as with the lexical diversity measure, trends within each group and divergence between the two groups were not large

Table 7 Summary of the lexical-diversity model. Degrees of freedom and *p*-values based on the Satterthwaite approximation. Limited-exposure group set as baseline

Random Effects			Variance				*SD*
Participant (Intercept)			0.001				0.023
Residual			0.000				0.016
Fixed effects	Estimate	Std. Error	*df*	*t* value	Pr(>|t|)	95% CI lower	95% CI upper (>|t|)
(Intercept)	0.439	0.007	80.15	60.82	<0.01	0.426	0.453
Exposure consistent	0.004	0.010	80.15	0.37	0.71	–0.015	0.022
Time in-sojourn 3	0.017	0.008	124.80	2.03	0.04	0.001	0.032
Time postsojourn 1 (5 months)	–0.001	0.008	123.90	–0.15	0.89	–0.017	0.014
Time postsojourn 2 (9 months)	–0.004	0.008	123.90	–0.46	0.64	–0.018	0.011
Time postsojourn 3 (4 years)	0.009	0.008	123.90	1.13	0.26	–0.006	0.023
Time postsojourn 4 (7 years)	0.002	0.007	123.90	0.27	0.79	–0.013	0.017
Exposure consistent × Time in-sojourn 3	0.005	0.011	124.50	0.51	0.61	–0.015	0.025
Exposure consistent × Time postsojourn 1	–0.006	0.011	124.50	–0.58	0.56	–0.026	0.014
Exposure consistent × Time postsojourn 2	–0.009	0.010	124.10	–0.90	0.37	–0.029	0.01
Exposure consistent × Time postsojourn 3	0.017	0.010	124.60	1.59	0.12	–0.003	0.037
Exposure consistent × Time postsojourn 4	0.001	0.010	124.40	0.08	0.93	–0.019	0.021

or consistent enough to attribute confidently to any particular time interval, even though the two groups ended up differently at postsojourn 4. A regression model of group differences over time is summarized in Table 9.

Table 8 Exposure and Vocabulary Knowledge Test descriptives

Condition	Statistic	Presojourn	In-sojourn 3	Postsojourn 3 (4 years)[a]	Postsojourn 4 (7 years)
Limited	*M*	2,854	3,183	3,091	3,171
	n	12	9	11	12
	SD	520	491	474	612
	95% CI Upper	3,148	3,504	3,371	3,517
	95% CI Lower	2,560	2,862	2,811	2,825
Consistent	*M*	2,784	3,372	3,667	3,750
	n	16	16	12	16
	SD	824	557	573	589
	95% CI Upper	3,188	3,645	3,991	4,038
	95% CI Lower	2,380	3,099	3,343	3,462

Note. [a]Time period refers to the amount of time after residence abroad ended.

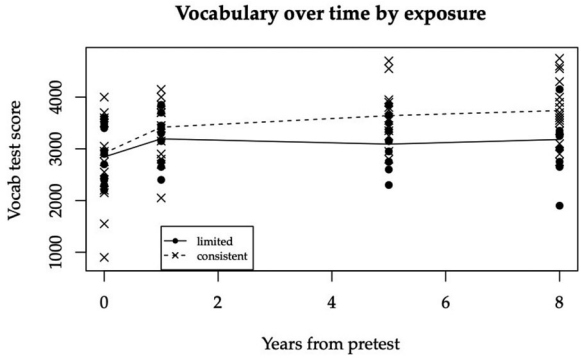

Figure 4 Exposure and Vocabulary Knowledge Test over time, split by group according to amount of continuing L2 exposure. Points are individuals' scores. Lines are locally weighted regressions showing trends in each of the two groups.

Long-Term Retention, Peak Instructional Attainment, and Exposure

The preceding analyses found significant effects of continuing exposure on EIT proficiency and speech rate. That is, the consistent-exposure group tended to perform higher on both measures at postsojourn 1 than the limited-exposure group. The question thus arises whether differences between the groups are really based on continuing exposure or simply reflect the fact that speakers who reached higher peak instructional attainment were more likely on average both to continue using the target language frequently and to score highly on tests at a later date. Figure 5 examines in more detail the relationship between EIT score at time point postsojourn 1 and EIT score at time point postsojourn 4, representing the end of the study.

Table 9 Summary of vocabulary model. Degrees of freedom and *p*-values based on the Sattherthwaite approximation. Limited-exposure group set as baseline

Random Effects					Variance		*SD*
Participant (Intercept)					173,094		416
Time postsojourn 3 x participant					216,421		465.2
Residual					223,036		472.3
Fixed effects	Estimate	Std. Error	*df*	*t* value	Pr(>|t|)	95% CI lower	95% CI upper
(Intercept)	2,854	182	44.7	15.7	<0.01	2,505	3,204
Exposure consistent	–70	240	44.7	–0.3	0.77	–532	393
Time in-sojourn 3	262	215	52.8	1.2	0.23	–148	678
Time postsojourn 3 (4 years)	–22	257	56.7	–0.1	0.93	–516	472
Time postsojourn 4 (7 years)	77	199	49.5	0.4	0.70	–303	458
Exposure consistent × Time postsojourn 3	326	272	50.8	1.2	0.24	–200	845
Exposure consistent × Time postsojourn 3	320	337	56.8	0.9	0.35	–330	963
Exposure consistent × Time postsojourn 4	3	271	51.5	0.01	0.99	–518	522

There is a substantial positive correlation between earlier EIT scores and later EIT scores for both groups. The slope of the relationship appears to be steeper for the group with limited continuing exposure. And participants with limited continuing exposure were more likely to show a decrease in EIT score over the course of the study than those with more consistent exposure, as shown by the number of points above and below the y = x diagonal. To examine whether differences in outcomes in the two groups are fully accounted for by the preexisting differences in EIT scores at time point postsojourn 1, we constructed a linear regression model of final EIT scores at postsojourn 4, with group scores at postsojourn 1 and their interaction as predictors. Because we are modeling only one datapoint per subject here, there is no random effects structure, and we use a plain general linear model. Results are summarized in Table 10.

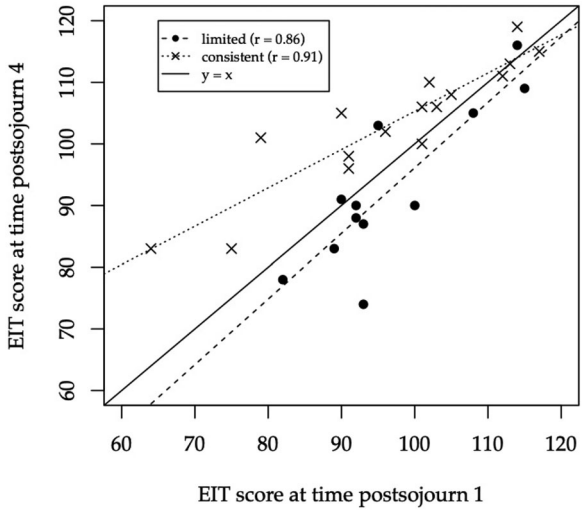

Figure 5 Elicited Imitation Test scores at time points postsojourn 1 and postsojourn 4. Points are individuals' scores. Solid line is the y = x diagonal; dashed and dotted lines are from simple regressions fit to the limited- and consistent-exposure groups, respectively.

Table 10 Summary of Elicited Imitation Test model. Limited-exposure group set as baseline

Effects	β	Std. Error	t value	Pr(>|t|)	95% CI lower	95% CI upper
(Intercept)	92.96	1.59	58.4	<0.001	89.84	96.08
Exposure consistent	10.48	2.11	5	<0.001	6.36	14.61
Postsojourn 1 score	1.06	0.16	6.6	<0.001	0.75	1.38
Exposure consistent × Post sojourn 1 score	–0.44	0.19	–2.4	0.027	–0.81	–0.07

The model shows that for every point higher a participant in the limited-exposure group scored on the EIT at postsojourn 1, they scored on average 1.06, 95% CI [0.75, 1.38], points higher at postsojourn 4. The slope of that relationship was 0.07–0.81 points smaller for participants in the consistent-exposure group, and the difference between these slopes was significant, as shown in the interaction. In other words, the relationship between peak instructional attainment and proficiency at postsojourn 4 was moderated by exposure. Independently of peak instructional attainment, participants with more consis-

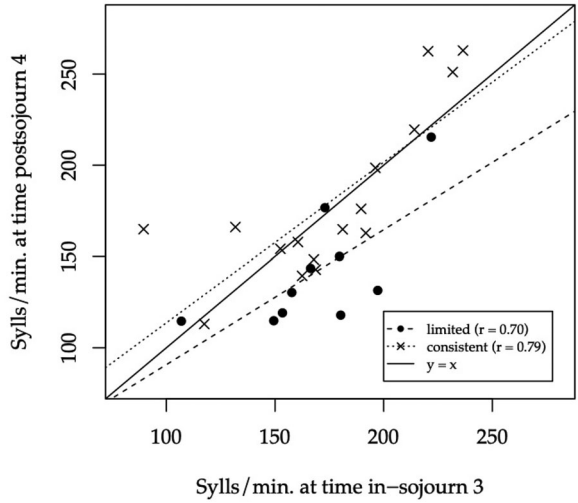

Figure 6 Fluency at time-points in-sojourn 3 and postsojourn 4. Points are individuals' scores. Solid line is the y = x diagonal; dashed and dotted lines are from simple regressions fit to the limited- and consistent-exposure groups, respectively.

tent exposure scored about 6.4–14.6 points higher on the EIT at the end of the study. This shows that their higher scores relative to the limited-exposure group were not simply a function of how proficient they became in the earlier stages of the study; consistent exposure was linked to higher proficiency 6 years after graduation (and 7 years after study abroad) regardless of peak instructional attainment. What the interaction showed is that, even though most of the variance in outcomes at postsojourn 4 within each group was predictable from the variance in their peak instructional attainment, the size of the predictive effect was larger for the group with less consistent exposure. In other words, the best way to end up a highly proficient speaker is to start out as a highly proficient speaker after your time abroad. But the benefit of being a highly proficient speaker at that earlier time, or equivalently the disadvantage of being a relatively nonproficient speaker at that earlier time, is not as large if one has continuing exposure to the language. One interpretation is that continuing exposure allows everyone to improve over time, reducing the differences between those who attained higher peak instructional attainment and those attaining lower levels.

Fluency

Figure 6 shows the corresponding plot for speech rate in the oral interview. As with the EIT, participants who reached higher levels by the end of study

Table 11 Summary of fluency model. Limited-exposure group set as baseline

Effects	β	Std. Error	t value	Pr(>|t|)
(Intercept)	145	9	16.4	<0.001
Exposure consistent	33	11	3	<0.01
In-sojourn 3 score	0.74	0.3	2.5	0.02
Exposure consistent × In-sojourn 3 score	0.14	0.35	0.4	0.69

abroad tended to still display higher levels 7 years later. As with the EIT, variance in peak instructional attainment explained most of the variance in speech rate at postsojourn 4 (though not quite as much as for the EIT). And as with the EIT, subjects with consistent exposure appeared to be more likely to improve over time than those with limited exposure. Unlike the EIT, however, there was no robust difference between groups in the size of the peak instructional attainment effect here, and the difference in slopes was minimal, as shown in Table 11. This does not provide support, then, for the idea that continuing exposure has a "leveling" effect on speech rate as it appears to for proficiency. Instead, peak attainment appears to be roughly as important for the group with consistent continuing exposure as it is for the group with more limited exposure. The relationship between peak instructional speech rate and postsojourn 4 speech rate does not appear to be moderated by exposure.

We did not examine this type of moderation model for the other two dependent variables, lexical diversity and vocabulary knowledge. This is because neither of these variables showed much difference over time after the sojourn, nor much difference between groups. As such, it was not appropriate to explore whether differences between groups were caused by exposure effects or preexisting peak instructional attainment differences.

Discussion

The goal of the current study was to explore several variables that may contribute to long-term retention of L2 proficiency. As Schmid (2023) emphasized,

> It is no exaggeration to say that we currently have no understanding of how—or even if—FL skills can attrite; of which grammatical or lexical features are more or less vulnerable, and why; of whether and how different subskills or modalities are differentially affected; of what other factors (length of time, amount of contact, attitudes) will facilitate or impede attrition and to what degree; nor in how former learners can be supported in maintaining or regaining proficiency and whether

> pedagogical approaches geared towards teaching a language the first time round are fit for purpose in re-learning.
>
> (p. 2)

The benefit of the data utilized in the current study is that they span 8 years (2011–2019), including the time period that Bahrick (1984) demonstrated to be important for attrition, the first 6 years after instruction. Data collection for the current study began presojourn, when participants were finishing their second year of university. Data were then collected three times while they were abroad and twice while they were completing their final year of university. Three and 6 years after graduation, we met with them again. No longer students, their lives were quite different from each other, as were their relationships with and use of their L2s. Due to the mixed-methods approach adopted in this study, we could utilize qualitative data to categorize participants as either having consistent or limited exposure to their L2s. Furthermore, the longitudinal quantitative data allowed us to document participants' peak instructional attainment, which for many of them was at the end of their year abroad. However, over half of our long-term participants (57%) have continued to use their L2 skills (not in formal instruction though), and some are still showing signs of development in different areas of their linguistic knowledge and use. Due to the different language assessments administered in our study (an oral proficiency test, an oral interview, and a vocabulary test), we could compare across subskills to begin to investigate many of the areas mentioned in Schmid (2023).

Results of research question 1, first of all, demonstrate very little attrition among our participants. That being said, we did find different patterns among our findings for general proficiency, oral fluency, lexical diversity, and vocabulary knowledge. Although on the oral proficiency test (i.e., EIT) we found a significant difference between the exposure groups at the 2019 posttest (postsojourn 4), the limited-exposure group did not show any significant change since returning from study abroad in 2012 (postsojourn 1) and continued to score at a level much higher than when the study began in 2011. The difference between the groups at postsojourn 4 was thus due to the consistent group continuing to show improvement over time and not to the limited-exposure group showing attrition. Regarding oral fluency, operationalized in the current study as speech rate in an oral interview, the results differed slightly. Both groups showed a decline after study abroad in their speech rate, but the consistent group improved again at postsojourn 3, whereas the limited group showed a significant decline at postsojourn 3. There is little change in the limited group's scores at postsojourn 4, however, suggesting that their fluency

may have stabilized between postsojourn 3 and postsojourn 4 (3–6 years after graduation).

Showing yet another pattern, lexical diversity in the oral interview, operationalized as MATTR, did not show significant differences between the groups at any of the time points. The consistent-exposure group did perform higher at postsojourn 3 and postsojourn 4, but the differences were not significant. For vocabulary knowledge (results of the X-lex test), the consistent-exposure group had somewhat higher scores at the end of the study, yet the differences between groups were not large or robust enough to be tracked with certainty to a particular interval of time. Mickan et al. (2023) found that language use played a role in retention of Spanish vocabulary as measured on a picture-naming test. Considering our results as well, it seems that exposure/use is important for retention when receptive vocabulary knowledge is measured using a controlled test. However, when productive vocabulary knowledge is measured on a more open task, such as an oral interview, the results are less clear. It may be the case that other measures of lexical richness, such as lexical sophistication, would be more sensitive to changes based on exposure.

In sum, the results of research question 1 demonstrate that participants generally retained the proficiency developed during their earlier language studies, regardless of subsequent level of exposure; only speech rate fluency showed a significant decline for the limited exposure group. However, the consistent-exposure group continued to improve to some degree, performing better than the limited-exposure group at the end of the study (postsojourn 4) in oral proficiency, oral fluency, and vocabulary (though not lexical diversity); and it was at postsojourn 3 (4 years after study abroad and 3 years after graduating) when these divergences between the groups became evident. This time period was also important in Bahrick (1984). Although the assessments used in the two studies were notably different, the findings suggest that an initial period of attrition is likely within the first 6 years postinstruction. However, the results also suggest that not all aspects of linguistic knowledge or performance change during this time, which highlights the importance of using multiple assessment types. In Bahrick's study, learners' abilities mostly stabilized after 6 years postinstruction. Our last data collection wave also took place 6 years after participants graduated. To test whether their abilities stabilized after that time, another data collection wave would be needed.

Research question 2 examined the relationship between long-term retention, peak instructional attainment, and exposure to explore the extent to which differences between exposure groups on the EIT and fluency in the oral interview at postsojourn 4 could reflect initial differences in peak instructional

attainment. In other words, could it be the case that speakers who reached higher peak instructional attainment at the end of their studies were more likely on average both to continue using the target language frequently and to score highly on tests at a later date? Overall, results of research question 2 demonstrate that improved outcomes on proficiency and fluency for the consistent group relative to the limited group are not entirely explained by peak instructional attainment; that is, consistent exposure per se also improves long-term outcomes. This leads to the question of whether peak instructional attainment works differently for the consistent and limited groups. Results for this question demonstrate that the variable of peak instructional attainment, as measured in the EIT, is relevant for long-term retention, but it is more important for the limited-use group compared to the consistent group. Participants in the limited group, who had higher peak instructional attainment at the end of their studies, attrited less. However, it is important to keep in mind that the amount of decline across all attainment levels was relatively small. Regarding fluency, the results differ somewhat from proficiency. For fluency, peak instructional attainment is also relevant for long-term retention, but it is equally important for both exposure groups. It thus appears that the effect of peak instructional attainment may work differently for the two groups. This finding is novel and not based on a particularly large sample of speakers, so we interpret it with caution until it can be replicated in other studies. Very few studies on L2 attrition/long-term retention have included measures of oral fluency, but results of the current study suggest that it is an important measure to include.

Results of this study demonstrate that consistent practice after completing a degree in a foreign language is a clear way to maintain advanced skills already acquired and also potentially to continue to improve. One of the challenges of measuring exposure in a longitudinal study such as this is how to do it reliably. Because some data points span 3 years, we had to find a way to categorize participants based on their experience over that time. We did not regularly ask them to answer questionnaires about language contact. Instead, in an oral interview in their L2 and another one in their L1, we asked them to tell us about their language use over the past 3 years. Participants often answered with much shame that they had not been practicing their French or Spanish and that they felt guilty about that. Even within the consistent group we witnessed much individual variation, from a participant who was living in Mexico and married to a Mexican citizen to a participant living in the UK, whose main exposure was the Spanish-language soap operas she watched nearly every day. To better understand how much practice is needed to maintain L2 skills, we need additional studies that investigate the issue of quantity and quality of exposure/rehearsal

in a more fine-grained way. Mickan et al. (2023) operationalized quality of Spanish input as native versus nonnative speaker interlocutors and found that speaking with native Spanish speakers was more beneficial for retention. Our method of categorizing participants into limited- or consistent-exposure groups considered interaction with other users of their L2 as well as other sources of input like digital media, and others. Although our method of operationalizing exposure could certainly be improved, we suggest that when considering quality of the input, other variables besides L1/L2 status should be considered.

One of the main findings of this study is the role that peak instructional attainment plays in long-term retention/attrition of second languages acquired primarily in an instructional setting and including a year abroad. Participants in our limited-exposure group still showed variation in their performance on the EIT and speech rate on the interview, but in general, most could be classified as advanced speakers of their L2s at the time of their graduation from university. Additional research is needed with learners of lower proficiency levels to better understand at what point learners might hit a general critical threshold of proficiency, as Neisser (1984) suggested, that supports long-term retention.

Conclusion

We posed the following questions at the outset of this manuscript: What happens to hard-earned language skills when students stop using them regularly? Do they change at all? Are they lost? Are some skills impacted more than others? The findings from our longitudinal data provide some indication that these skills are not lost after 6 years when an advanced level of proficiency is reached but that some skills are impacted more than others. In particular, oral fluency benefits from continued use. We also asked, what activities help in the long-term maintenance of advanced language skills? Based on the participants in the current study, continued exposure to the target language not only appears to protect against attrition, but for some skills, even results in continued improvement after instruction ends.

This study represents only some of what is possible using the LANGSNAP corpus. As nearly all of the longitudinal data used in this study are publicly available (only data which cannot be fully anonymized are unavailable), this allows for other researchers to begin to utilize this resource to explore some of the issues raised by Schmid (2023).

Final revised version accepted 27 April 2024

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at https://osf.io/sypbn/ and https://osf.io/g64nh/. All proprietary materials have been precisely identified in the manuscript.

Note

1 Advanced proficiency has multiple definitions in the literature (e.g., Byrnes, 2012; Olson, 2023). We consider the learners in the current study as advanced given the following evidence: Learners in the current study were pursuing Bachelor's degrees in French and/or Spanish, were required to write an undergraduate dissertation in the target language, and had been studying the target language for 8.4 years (SD = 4.1). Their oral proficiency was also assessed independently as part of the current study and demonstrated that upon returning from their year abroad, they had achieved a level representative of advanced proficiency using Bowden (2016)'s benchmarks: Advanced M = 96.5 (range: 61–116); Very advanced M = 109 (range: 99–119); learners in the current study fall within these ranges in each of the postsojourn periods (see Table 2).

References

Bahrick, B. P. (1984). Fifty years of second language attrition: Implications for programmatic research. *The Modern Language Journal*, *68*(2), 105–118. https://doi.org/10.1111/j.1540-4781.1984.tb01551.x

Bardovi-Harlig, K., & Stringer, D. (2010). Variables in second language attrition: Advancing the state of the art. *Studies in Second Language Acquisition*, *32*(1), 1–45. https://doi.org/10.1017/S0272263109990246

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015b). Parsimonious mixed models. arXiv:1506.0496z

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015a). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bowden, H. W. (2016). Assessing second language oral proficiency for research. *Studies in Second Language Acquisition*, *38*(4), 647–675. https://doi.org/10.1017/S0272263115000443

Byrnes, H. (2012). Of frameworks and the goals of collegiate foreign language education: Critical reflections. *Applied Linguistics Review*, *3*(1), 1–24. https://doi.org/10.1515/applirev-2012-0001

Covington, M. (2007). *MATTR user manual*. https://ai1.ai.uga.edu/caspr/MATTR-Manual.pdf.

De Bot, K., & Clyne, M. (1989). Language reversion revisited. *Studies in Second Language Acquisition*, *11*(2), 167–177. https://doi.org/10.1017/S0272263100000590

Fergadiotis, G., Wright, H. H., & Green, S. B. (2015). Psychometric evaluation of lexical diversity indices: Assessing length effects. *Journal of Speech, Language, and Hearing Research*, *58*(3), 840–852. https://doi.org/10.1044/2015_JSLHR-L-14-0280

Flores, C. (2010). The effect of age on language attrition: Evidence from bilingual returnees. *Bilingualism: Language and Cognition*, *13*(4), 533–546. https://doi.org/10.1017/S136672890999054X

Flores, C. (2015). Losing a language in childhood: A longitudinal case study on language attrition. *Journal of Child Language*, *42*(3), 562–590. https://doi.org/10.1017/S0305000914000233

Gelman, A., & Stern, H. (2006). The difference between significant and non-significant is itself not statistically significant. *The American Statistician*, *60*(4), 328–331. https://doi.org/10.1198/000313006×152649

Hansen, L. (1999). Not a total loss: The attrition of Japanese negation over three decades. In L. Hansen (Ed.), *Second language attrition in Japanese contexts* (pp. 142–153). Oxford University Press.

Hansen, L. & Chen, Y.-L. (2001). What counts in the acquisition and attrition of numeral classifiers? *JALT Journal*, *23*(1), 90–110.

Huensch, A., & Tracy-Ventura, N. (2017). L2 utterance fluency development before, during, and after residence abroad: A multidimensional investigation. *Modern Language Journal*, *101*(2), 275–293. https://doi.org/10.1111/modl.12395

Kinginger, C., & Zhuang, J. (2023). *The professional lives of language study abroad alumni: A mixed methods investigation*. Multilingual Matters.

Kuhberg, H. (1992). Longitudinal L2-attrition versus L2-acquisition, in three Turkish children-empirical findings. *Second Language Research*, *8*(2), 138–154. https://doi.org/10.1177/026765839200800203

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lambert, R. D., & Freed, B. F. (1982). *The loss of language skills*. Newbury House Publishers.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.

Marsden, E., Mackey A., & Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In A. Mackey & E. Marsden (Eds.), *Advancing methodology and practice: The IRIS repository of instruments for research into second languages* (pp. 1–21). Routledge.

McManus, K., Mitchell, R., & Tracy-Ventura, N. (2014). Understanding insertion and integration in a study abroad context: The case of English-speaking sojourners in France. *Revue Française de Linguistique Appliquée*, *19*(2), 97–116. https://doi.org/10.3917/rfla.192.0097

Meara, P. M. (2005). *X_Lex: The Swansea vocabulary levels test* (Version 2.05) [Computer software]. Lognostics.

Mehotcheva, T. H. (2010). *After the fiesta is over: Foreign language attrition of Spanish in Dutch and German Erasmus students* [Doctoral dissertation, University of Groningen].

Mickan, A., McQueen, J., Brehm, L., & Lemhofer, K. (2023). Individual differences in foreign language attrition: A 6-month longitudinal investigation after study abroad. *Language, Cognition, and Neuroscience*, *38*(1), 11–39. https://doi.org/10.1080/23273798.2022.2074479

Mickan, A., McQueen, J. M., & Lemhöfer, K. (2020). Between-language competition as a driving force in foreign language attrition. *Cognition*, *198*, Article 104218. https://doi.org/10.1016/j.cognition.2020.104218

Mitchell, R., Tracy-Ventura, N., & Huensch, A. (2020). After study abroad: The maintenance of multilingual identity among Anglophone languages graduates. *Modern Language Journal*, *104*(2), 327–344. https://doi.org/10.1111/modl.12636

Mitchell, R., Tracy-Ventura, N., & McManus, K. (2017). *Anglophone students abroad: Identity, social relationships and language learning*. Routledge.

Murtagh, L. (2003). *Retention and attrition of Irish as a second language. A longitudinal study of general and communicative proficiency in Irish among second level school leavers and the influence of instructional background, language use and attitude/motivation variables* [Doctoral dissertation, Rijksuniversiteit Groningen].

Neisser, U. (1984). Interpreting Harry Bahrick's discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General*, *113*(1), 32–35. https://doi.org/10.1037/0096-3445.113.1.32

Olson, D. J. (2023). A systematic review of proficiency assessment methods in bilingualism research. *International Journal of Bilingualism*, *28*(2). https://doi.org/10.1177/13670069231153720

Ortega, L. (2000). *Understanding syntactic complexity: The measurement of change in the syntax of instructed L2 Spanish learners* [Doctoral dissertation, University of Hawaii].

Ortega, L., Iwashita, N., Norris, J. M., & Rabie, S. (2002, October). *An investigation of elicited imitation tasks in crosslinguistic SLA research*. Paper presented at the Second Language Research Forum, Toronto.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. John Benjamins.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. In B. Kopke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 121–133). John Benjamins.

Schmid, M. (2011). *Language attrition*. Cambridge University Press.
Schmid, M. (2023). The final frontier? Why we have been ignoring second language attrition, and why it is time we stopped. *Language Teaching*, *56*(1), 73–93. https://doi.org/10.1017/S0261444822000301
Schmid, M., & Mehotcheva, T. (2012). Foreign language attrition. *Dutch Journal of Applied Linguistics*, *1*(1), 102–124. https://doi.org/10.1075/dujal.1.1.08sch
Schmid, M. S., & Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, *7*(6), 637–667. https://doi.org/10.1075/lab.17058.sch
Thomas, M. (2006). Research synthesis and historiography: The case of assessment of second language proficiency. In J. M. Norris & L. Ortega (Eds.) *Synthesizing research on language learning and teaching* (pp. 279–298). John Benjamins.
Tomiyama, M. (2000). Child second language attrition: A longitudinal case study. *Applied Linguistics*, *21*(3), 304–332. https://doi.org/10.1093/applin/21.3.304
Tomiyama, M. (2008). Age and proficiency in L2 attrition: Data from two siblings. *Applied Linguistics*, *30*(2), 253–275. https://doi.org/10.1093/applin/amn038
Tracy-Ventura, N., McManus, K., Norris, J., & Ortega, L. (2014). "Repeat as much as you can": Elicited imitation as a measure of oral proficiency in L2 French. In P. Leclerq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 143–166). Multilingual Matters.
Weltens, B. (1989). *The attrition of French as a foreign language*. Foris Publications.

Supporting Information

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

Acessible Summary