



Experiential, perceptual, and cognitive individual differences in the development of declarative and automatized phonological vocabulary knowledge

Research Article

Kazuya Saito¹ and Takumi Uchihara²

Cite this article: Saito, K., & Uchihara, T. (2024). Experiential, perceptual, and cognitive individual differences in the development of declarative and automatized phonological vocabulary knowledge. *Bilingualism: Language and Cognition*, 1–17. <https://doi.org/10.1017/S1366728924000609>

Received: 29 September 2023
Revised: 7 June 2024
Accepted: 9 June 2024

Keywords:

listening; spoken vocabulary; individual differences; automatization; aptitude effects; experience effects

Corresponding author:

Kazuya Saito;
Email: k.saito@ucl.ac.uk

¹Institute of Education, University College London, London, UK and ²Graduate School of International Cultural Studies, Tohoku University, Sendai, Miyagi, Japan

Abstract

The present study explores the influence of individual differences in experience, perceptual acuity, and working memory on the development of both declarative and automatized aspects of L2 phonological vocabulary knowledge. A total of 486 Japanese English-as-a-foreign-language (EFL) students took part in two vocabulary tests designed to measure declarative (meaning recognition) and automatized knowledge (lexicosemantic judgement task). Their performance was tied to the quantity and quality of their EFL experience, as well as their scores in auditory processing and working memory. While several significant, modest correlations between experience, aptitude, and vocabulary outcomes were observed, certain predictor variables were *uniquely* associated with either declarative or automatized vocabulary performance. Specifically, individuals with more extensive, typically language-focused EFL training and greater working memory demonstrated higher levels of declarative knowledge. Conversely, those who pursued extracurricular practice outside the classroom – exposing themselves to auditory materials and/or participating in study-abroad experiences – showed a more automatic execution of vocabulary knowledge.

1. Introduction

In our increasingly globalized world, many adult second language (L2) learners pursue foreign language education in classroom settings. However, the learning outcomes in these settings are marked by significant individual variation. This is because every learner’s profile is unique, influenced by an array of contextual factors. These factors include the age of learning (Larson-Hall, 2008), length of learning (Jaekel et al., 2017), and engagement in extracurricular activities (e.g., Muñoz, 2014 for study abroad; Saito & Hanzawa, 2016 for conversation activities). Research into these determinants holds considerable pedagogical value, as the insights can elucidate the factors that most effectively facilitate successful classroom-based L2 learning.

To evaluate learning outcomes, the skill acquisition theory for instructed L2 acquisition (DeKeyser, 2017; Suzuki, 2023) differentiates between two types of knowledge. The first, declarative knowledge, encompasses the concept of “knowing what.” Typically operationalized as a learner’s metalinguistic understanding (e.g., grammar rules, vocabulary form and meaning), this type of knowledge is usually taught through explicit instruction and evaluated using language-focused testing modalities (e.g., multiple choice, fill-in-the-blanks). Such tests allow learners to access this knowledge consciously. On the other hand, automatized knowledge relates to “knowing how” in *real-life* language use contexts. After acquiring explicit declarative knowledge, learners transition to proceduralizing it initially through controlled tasks that minimize communicative pressure (e.g., grammar drills, vocabulary flashcards). Over time, with repetitive practice and sustained exposure, learners can swiftly and reliably access this knowledge without much conscious thought. It is often considered the ultimate goal of L2 learning, especially relevant for communicatively authentic listening and speaking scenarios.

In their comprehensive overview, Suzuki and Elgort (2023) pointed out while much has been documented about L2 morphosyntax acquisition, other areas of language have been somewhat ignored. Notably, there is an emerging paradigm surrounding the assessment of the most critical skill for L2 speech comprehension – i.e., phonological vocabulary (McLean et al., 2015; Saito et al., 2023; Uchihara et al., 2024). Within the context of 486 Japanese English-as-a-Foreign-Language (EFL) learners spanning varied background and proficiency levels, the present study seeks to determine the extent to which adult L2 learners can develop declarative and automatized phonological vocabulary knowledge after extensive amount of

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



This article has earned badges for transparent research practices: Open Materials. For details see the Data Availability Statement and Supplementary Information.

classroom L2 learning, and then identify which individual difference factors related to experience (quantity and quality of EFL experience) and aptitude (auditory processing and working memory) can lead to such outcomes.

2. Background

2.1. Declarative and automatized phonological vocabulary knowledge

Few disagree that the development of advanced L2 listening comprehension is a catalyst for successful communication in social, business, and academic settings. L2 listening comprehension requires multiple linguistic abilities. After L2 learners break down auditory information into words (Norris & McQueen, 2008), they need to discern the distinct morphological attributes of each word and comprehend the grammatical patterns in a sentence (Vafae & Suzuki, 2020). Moreover, L2 learners need to understand a speaker's intention in light of conversational, societal, and cultural backdrops (Taguchi, 2011) and extralinguistic factors (e.g., vocal tone, facial cues, and bodily gestures; Kamiya, 2022). To date, numerous studies have examined which linguistic abilities are relatively important for global L2 listening proficiency, and they have consistently shown that vocabulary knowledge accounts for by far the largest amounts of variances in L2 learners' listening test scores ($r = .5-.7$; see Zhang & Zhang, 2022 for a meta-analysis) and that vocabulary knowledge thus needs to be prioritized in an L2 syllabus (e.g., Vafae & Suzuki, 2020; Vandergrift & Baker, 2018; Wallace, 2022).

According to Nation's (2013) oft-cited model, word knowledge relevant to successful L2 listening can be defined as the ability to understand not only what target words sound like and mean (i.e., form-meaning mapping), but also how they interact with other words in a semantically, collocationally, and grammatically appropriate manner (i.e., use-in-context). While the former deals with individual word recognition, the latter emphasizes the ability to understand these words within broader contexts. This requires the processing of morphosyntax, pragmatics, and paralinguistic cues to ensure effective L2 listening. Nation's two-step framework for phonological vocabulary knowledge here corresponds to the skill acquisition theory which stresses the distinction between declarative knowledge (the association between word forms and meanings) and automatized knowledge (the quick and stable recognition of words in relation to neighboring words in sentences).

As for the form-meaning mapping stage of phonological vocabulary knowledge, increasing evidence suggests that many EFL learners can recognize words when they see them in writing, but struggle to recognize the same words when they are audibly presented without orthographic cues (Cheng & Matthews, 2018; Hamada & Yanagawa, 2023; Milton & Hopkins, 2006). This challenge may arise because numerous EFL learners predominantly engage with written materials and exercises, often lacking exposure to authentic auditory input in their target language (Muñoz, 2014). Reports frequently highlight that few L2 learners receive adequate phonetic training or possess a robust phonological awareness of the L2 system (see Saito, 2019 for the case of Japanese EFL learners). Thus, scholars have begun to suggest that the assessment of phonological vocabulary knowledge needs to involve both auditory and written modalities to capture the nature of the form-meaning aspects of the knowledge in L2 listening. To this end, for example, scholars have adopted the audio versions of multiple choice meaning recognition (MR) (McLean et al., 2015), meaning recall (Cheng et al., 2022), and

yes/no form recognition (Milton & Hopkins, 2006). Some scholars have explored measuring reaction time during single word recognition tasks (e.g., Hui & Godfroid, 2021).

Schmitt (2019) highlighted that limited research has delved into the use-in-context dimensions of vocabulary knowledge. The lack of research is possibly due to the intricate nature of word usage in L2 listening comprehension, where understanding requires learners to grasp the semantic, morphosyntactic, and phraseological relationship between target words and their neighbors. To explain the complex nature of spoken word recognition in real-life listening contexts, N. Ellis (2006) characterized humans as "optimal word processors" (p. 2). Upon receiving auditory input even at sub-lexical levels, listeners can swiftly discern potential word matches, using context to hone in on the most likely candidates. This contextual processing is attuned to frequency such that words that commonly co-occur with surrounding words are given precedence. Indeed, native listeners can recognize word combinations with higher mutual information scores (indicating stronger meaning associations) more rapidly than their L2 counterparts (N. Ellis et al., 2008).

In the realm of L2 morphosyntax learning, as opposed to metalinguistic tests that gauge declarative knowledge (e.g., multiple-choice and fill-in-the-blanks; R. Ellis, 2005), grammaticality judgment tasks (GJTs) have often been used to assess the degree of learners' automatized knowledge. In GJTs, learners listen to sentences featuring manipulated target morphosyntactic structures and must quickly decide whether each sentence is grammatically correct or incorrect (see Plonsky et al., 2020 for a review). Ample evidence suggests that GJT scores fundamentally differ from those of metalinguistic tests, indicating that the former captures automatized knowledge while the latter reflects declarative knowledge (Gutiérrez, 2013). Furthermore, performance on GJTs has been shown to correlate with key variables affecting the attainment of high-level L2 proficiency, such as age of acquisition (e.g., Abrahamsson & Hyltenstam, 2009) and length of immersion (e.g., Faretta-Stutenberg & Morgan-Short, 2018).

Building on the methodological paradigm underlying GJTs, an approach has introduced the lexicosemantic judgement task (LJT) to assess automatized phonological vocabulary knowledge (Saito et al., 2023; Uchihara et al., 2024). In the LJT, learners are presented with sentences that are grammatically correct and consist solely of high-frequency words. Despite this, the sentences are categorized based on their semantic use of a target word (e.g., "publish"). In some sentences, the target word is used appropriately (e.g., "He has published many books"), while in others, it is used in a semantically incongruous manner (e.g., "He has published many shoes"). Upon hearing each sentence and without much time to deliberate on the correct or incorrect use of the word, learners are asked to judge its semantic appropriateness.

In our prior projects (Saito et al., 2023; Uchihara et al., 2024), we explored the differences and degrees of variation in L2 learners' vocabulary knowledge when evaluated through two tasks focusing on form-meaning mappings (MR and recall) and one task centered on automatization (LJT). The findings revealed that participants' scores on the LJT were distinguished as different latent variables compared to their scores on MR and recall. Furthermore, the LJT scores showed a stronger correlation with global L2 listening proficiency ($r = .6-.7$) than did the scores for MR and recall ($r = .4-.5$).

Given that the prior studies have examined the measurements of two distinct facets of phonological vocabulary knowledge (i.e., MR and recall for declarative knowledge and LJT for

automatized knowledge), this raises pertinent questions: To what extent can L2 learners acquire these different dimensions of L2 phonological vocabulary knowledge? What factors contribute to successful spoken word learning? And how can pedagogy be adjusted to assist L2 learners in achieving automatized phonological vocabulary knowledge? As stated in the skill acquisition theory of instructed SLA, learning is conceptualized as the transition from declarative to procedural knowledge through consistent practice, ultimately leading to automatization.

The subsequent sections will review various factors identified by scholars as influential in the acquisition of phonological vocabulary knowledge. Consistent with the skill acquisition theory for instructed second language acquisition (DeKeyser, 2017; Suzuki, 2023), learning outcomes can be attributed to a combination of learner-external and learner-internal factors. The first category encompasses the quantity and quality of EFL experience – specifically, the extent and manner in which L2 learners have practiced the language both inside and outside the classroom. The second category includes factors related to *perception* and *cognition* dimensions of aptitude, respectively – focusing on how effectively L2 learners perceive and process input, with a particular emphasis on auditory processing and working memory. While the literature discusses a variety of aptitude variables (for a comprehensive review, see Li, 2016), our study specifically concentrated on two domain-general abilities pertinent to L2 phonological vocabulary learning: auditory processing at the *lower-order/perceptual* level and working memory at the *higher-order/cognitive* level.

2.2. Experience factors affecting vocabulary knowledge development

Unlike naturalistic settings where learners are exposed to a target language daily, classroom environments often offer limited quantity and quality of input. In these settings, learners typically engage in form-oriented instruction for a few hours, focusing primarily on rote memorization of vocabulary and grammar drills. It has been shown that the outcomes of classroom L2 learning (typically measured via global listening and speaking tests) can be determined not only by the timing and length of L2 learning within classrooms (Jaekel et al., 2017; Larson-Hall, 2008), but also by whether, to what degree, and how L2 learners seek more practice opportunities outside of classrooms (Muñoz, 2014; Saito & Hanzawa, 2016).

Though limited, some scholars have examined the extent to which L2 learners develop their vocabulary knowledge after years of EFL education and what factors impact their learning outcomes. In terms of L2 learners' form-meaning aspects of vocabulary knowledge (measured via MR), earlier studies suggested that 500–1,000 hours of instruction are needed for the acquisition of 1,000–2,000 word families and 2,000–2,500 hours are needed for the acquisition of 3,000–4,000 word families (Schmitt, 2008; Webb & Chang, 2012; but see McLean et al., 2014). More recently, scholars have further examined how not only quantity but also quality of EFL experience impacts on the acquisition of vocabulary knowledge in Spanish L1 speakers in Spain (González Fernández & Schmitt, 2015; Muñoz, 2011), Dutch L1 speakers in Belgium (Peters et al., 2019), and Chinese L2 speakers in China (Lu & Dang, 2023).

The findings showed that the attainment of advanced L2 vocabulary size (5,000–10,000 word families) can be related to the length of EFL training, along with recent exposure to the

target language both inside and outside the classroom (Muñoz, 2011). In terms of quality of input, more advanced L2 vocabulary knowledge can be related to certain activities, such as browsing the internet, watching movies/TV without subtitles, but not to reading written materials (Peters et al., 2019; but see De Wilde et al., 2020). Similar results were also found when it comes to the acquisition of collocation knowledge (González Fernández & Schmitt, 2015). The experience-related variables do not seem to make any impact on the acquisition of high frequency words (1,000 word families; Lu & Dang, 2023).

It is worth noting that existing literature has predominantly focused on the declarative dimensions of vocabulary knowledge (measured via MR and recall). As indicated in prior research (e.g., Wallace, 2022), achieving a robust phonological form-meaning mapping of L2 words is essential for successful L2 listening comprehension. However, it does not necessarily ensure that learners can readily access this knowledge during real-life listening tasks, which often demand attention to various language aspects beyond vocabulary (Saito et al., 2023; Uchiyama et al., 2024). To date, we have yet to know how learners can automatize such phonological knowledge, which is arguably the ultimate goal for many L2 learners. The current study was designed to address these concerns.

2.3. Aptitude factors affecting vocabulary knowledge development

Even if two individuals with same age and motivation profiles learn an L2 in the same way spending the same amount of time, their learning outcomes could still substantially differ. One reason could be tied to learners' intrinsic differences in perceptual and cognitive abilities. At the domain-specific level, a range of test formats have been proposed and their test scores have been found to significantly predict the outcomes of L2 learning in classroom settings (e.g., Carroll & Sapon, 1959 for Modern Language Aptitude Test; Meara, 2005 for LLAMA). To further examine precisely which aspects of perceptual and cognitive abilities affect L2 learning, scholars have begun to focus on the complex relationships between a range of domain-general abilities in L2 learning (e.g., Hi-Lab for Linck et al., 2013). In the context of phonological vocabulary learning (the main focus of the current study), the two domain-general abilities have received a growing amount of attention, auditory processing and working memory, and there is accumulating evidence showing that these aptitude factors interact to affect the development of various L2 skills in foreign language classrooms with relatively medium effects (for a comprehensive review, see Wen & Skehan, 2021).

2.3.1. Auditory processing

Researchers have posited that individuals exhibit differences in their perceptual capabilities to discern basic sound characteristics, such as frequency, duration, and intensity (i.e., auditory processing) and that such individual variations at sensory levels can influence a myriad of developmental outcomes, including language acquisition. For instance, children's auditory profiles have been linked to the likelihood of language impairment (Goswami, 2015) and the speed of language development (Kalashnikova et al., 2019). Extending this paradigm to adult L2 learning, recent studies have found auditory processing to be a key determinant of various L2 learning outcomes, including both phonology (Kachlicka et al., 2019) and lexicogrammar (Saito et al., 2022). The influence of this relatively perceptual,

lower-order aptitude becomes particularly evident when analyzing L2 learners in immersion settings wherein they can access ample L2 input on a daily basis if actively sought but this contrasts with classroom settings wherein learners often encounter restricted amounts of communicatively genuine input (Saito et al., 2022).

2.3.2. Working memory

One widely researched domain-general aptitude variable is working memory. While a range of frameworks exist (Baddeley, 2000; Miyake et al., 2001), Li's (2016) research synthesis has shown that two components of working memory have received much attention in the field of cognitive psychology and L2 acquisition research. They include phonological short-term memory (the capacity to temporarily retain perceived auditory information) and executive function working memory (the ability to manipulate stored information). Individual variations in working memory have been associated with global L2 skills such as listening and reading (Linck et al., 2013) and speaking (O'Brien et al., 2007). On a more specific level, numerous studies have explored the influence of working memory on different facets of L2 morpho-syntax development, considering variables like the type of intervention received (e.g., intentional vs. incidental; Tagarelli et al., 2011) and the learning context (e.g., Faretta-Stutenberg & Morgan-Short, 2018). In relation to L2 vocabulary acquisition (i.e., the main focus of the current study), though the research is somewhat sparse, findings indicate that L2 learners with superior working memory tend to achieve better learning outcomes regardless of the learning conditions (e.g., Bisson et al., 2021; Elgort et al., 2018; Perez, 2020). Different from the perceptual aptitude (auditory processing), this relatively cognitive, higher-order aptitude can be associated with L2 learning in both classroom and naturalistic settings (Li, 2016).

3. Current study

In the context of 486 Japanese EFL learners with varied proficiency levels, the primary objective of the current project was to investigate the factors affecting the development of both declarative and automatized dimensions of phonological vocabulary knowledge. Our study is characterized not only by its relatively large sample size, which included participants with a wide spectrum of L2 proficiency levels, but also by its comprehensive analytical approach. We uniquely incorporated the examination of both learner-external (experience-related) and learner-internal (aptitude-related) factors. As such, we aimed to shed light on the underlying mechanisms which determine how individuals learn and attain advanced L2 phonological vocabulary knowledge in classroom settings. Two primary research questions were formulated, along with their respective predictions:

1. To what extent do adult L2 learners acquire phonological vocabulary knowledge after extensive years of foreign language education?
2. Which variables related to experience and perceptual-cognitive abilities influence the attainment of declarative and automatized phonological vocabulary knowledge?

For RQ1, we assessed participants' knowledge using MR (McLean et al., 2015) and LJT's (Uchihara et al., 2024). Since our interest lies in the automatization of phonological vocabulary, the project highlighted *experienced* EFL learners with at least six

years of EFL education (for similar decisions, see Muñoz, 2011, 2014). Given the tendency of EFL education to emphasize form-oriented lessons, prioritizing the development of declarative knowledge over proceduralization and automatization (Suzuki, 2023), we expect many participants to exhibit high performance in MR (reflecting declarative knowledge). However, we anticipate fewer participants to excel in lexicosemantic judgements (reflecting automatized knowledge).

For RQ2, our hypothesis posits differential relationships between individual differences in two different types of phonological vocabulary knowledge (declarative vs. automatized) and relevant experiential variables. The declarative aspect of this knowledge might correlate with the extent of English study in classrooms (Jaekel et al., 2017). In contrast, the automatized facet might be more influenced by extracurricular activities outside the classroom, wherein learners apply previously acquired knowledge in more global contexts, such as extensive reading (e.g., De Wilde et al., 2020), extensive listening and watching (Peters et al., 2019), and study abroad experiences (Muñoz, 2014). We posited two distinct hypotheses regarding the effects of aptitude. Given that auditory processing is more a perceptual skill than a cognitive ability, such lower-order perceptual-cognitive variations can act as a bottleneck for every facet of language learning (Goswami, 2015). As a result, they might influence both declarative and automatized phonological vocabulary development. On the other hand, individual differences in working memory enable learners to retain and further process information, which is directly associated with mapping form to meaning in linguistic knowledge (i.e., declarative phonological vocabulary development; Perez, 2020) whereas long-term memory capacities have been associated with the evolution of automatized linguistic knowledge (e.g., procedural memory; Faretta-Stutenberg & Morgan-Short, 2018).

3.1. Project setup

This study was initiated between 2021 and 2023 as part of a broader project with dual objectives. The first objective was to explore methodologies for measuring automatized (as opposed to declarative) phonological vocabulary knowledge among experienced Japanese EFL learners, which was the focus of our prior projects (Saito et al., 2023; Uchihara et al., 2024). The second, and primary objective of the current paper, was to identify the factors that affect the acquisition of both declarative and automatized phonological vocabulary knowledge.

During the data collection period, a total of 486 participants were recruited. The initial 240 participants were also involved in the two preceding projects, while the subsequent 246 were exclusively part of the current study. The data from all participants were analyzed in this paper, with a clear demarcation in content between the prior projects (i.e., the development and validation of the lexicosemantic judgment task as a measuring method of automatized phonological knowledge) and the current paper (i.e., the experiential, perceptual, and cognitive foundations of automatized phonological vocabulary knowledge).

At the outset of the project, there was a discernible gap in knowledge regarding the measurement of automatized phonological vocabulary, with the existing literature primarily focused on declarative knowledge measures, such as MR and recall. In response, the first phase of our project in 2021 aimed to develop, test, and refine one of the inaugural outcome measures for automatized phonological vocabulary knowledge. As outlined in

Uchihara et al. (2024), this phase involved 114 participants undertaking two form-meaning tests, an automatization test, and completing the EFL experience questionnaire along with aptitude tests.

The second phase in 2022 further explored the relationship between the declarative and automatized dimensions of phonological vocabulary knowledge and L2 listening proficiency. As detailed in Saito et al. (2023), this phase included 126 participants undergoing a series of tests and questionnaires, alongside a global listening proficiency test (TOEIC).

Following the validation of the methodological framework in the initial phases, we proceeded to the third phase of data collection for the remaining 246 participants, concluding on 31 July 2023. All participants in the study undertook two vocabulary tasks, along with the EFL experience questionnaire and aptitude tests evaluating auditory processing and working memory. While the previous projects detailed the relationships between lexicosemantic judgement and other vocabulary and listening test scores among the initial 240 participants, the current paper elucidates the experiential and aptitude-related factors influencing the attainment of automatized phonological vocabulary knowledge across the entire dataset.

3.2. Participants

The final dataset comprises a total of 486 participants (198 males, 288 females) who spanned various university-level programs, including both undergraduate and graduate courses across Japan (M age = 25.2 years; $Range$ = 19–36 years). Through electronic flyers, we reached out to over 20 universities, targeting students who had completed at least six years of EFL education up to the high school level. According to the JACET SLA (2013) guidelines, we anticipated that such students would fall within a CEFR level between A2 and B1, categorizing them as functional users of L2 English. As Yashima et al. (2004) highlighted, the typical Japanese EFL curriculum starts with a concentration on vocabulary memorization, idiom acquisition, and sentence translation exercises. Gradually, the focus shifts to spoken communication and conversation exercises. Given the participants' substantial EFL background, we surmised that they had dedicated sufficient time to proceduralizing their declarative vocabulary knowledge throughout their lengthy EFL education, culminating in a certain level of automatized vocabulary knowledge (cf. for the initial stages of vocabulary development, see Lu & Dang, 2023).

Due to COVID-related restrictions on face-to-face interactions in Japan, we shifted our data collection process online, utilizing the Gorilla psychology experiment builder (Anwyl-Irvine et al., 2020). To ensure the highest quality of online data, several specific measures were taken. While over 500 participants expressed interest, we were ultimately able to include 486 in the study and final analyses:

1. We advised participants to use headphones for clearer audio without external noise and to ensure a stable internet connection on their computers. This setup was crucial for accessing high-quality sounds in a distraction-free environment.
2. Participants began with a preliminary vocabulary test and a working memory test (both forward and backward digit span). The vocabulary test assessed participants' ability to recognize the meanings of 20 words selected from the first 1,000 word families available in the BNC-COCA corpus, using Cobb's Vocab Profilers (<https://www.lex tutor.ca/>). If their

accuracy was below 80%, they were excluded from the study, as it was assumed they lacked the essential vocabulary knowledge necessary for minimal comprehension of L2 discourse (a similar criterion was applied in Dang et al., 2022). Participants who were unable to type and record their responses in the working memory task were also excluded, as they did not possess the necessary computer skills for participating in online experiments.

3. Once we confirmed that participants had successfully completed the preliminary tests, we provided them with a detailed instruction handout in Japanese to aid their comprehension of the tasks.
4. Then, a research assistant arranged a videoconference meeting with groups of 10–20 participants. During the meeting, the research assistant explained the tasks and their procedures, followed by a Q&A session.
5. Finally, we sent them a URL link that granted access to the main phonological vocabulary tests (MR and lexicosemantic judgment). Participants completed these tasks at their own pace using their computers. They had the option to seek guidance from the research assistant whenever they had questions.

The total duration of the data collection lasted for about one hour and was operationalized in the following order: LJT (15–20 minutes), MR (10–15 minutes), auditory processing (10–15 minutes), and EFL questionnaire (5–10 minutes). Instructions for all aspects of data collection, including vocabulary tests and EFL experience questionnaires, were provided in Japanese. Participants were encouraged to reach out to research assistants, who were native Japanese speakers, should they have had any inquiries regarding the procedure. Throughout the project, no participants reported the lack of their understanding of the task instructions in Japanese.

3.3. Phonological vocabulary measures

Following the prior projects (Saito et al., 2023; Uchihara et al., 2024), two test formats, MR and LJT, were adopted to measure the declarative (form-meaning) and automatized (use-in-context) dimensions of phonological vocabulary. All test resources can be found on the open science platform, L2 Speech Tools, as detailed by Mora-Plaza et al., 2022 at <http://sla-speech-tools.com/>. Demo versions of the tests can be accessed via <https://app.gorilla.sc/openmaterials/663422> (see Supporting Information-S1). Note that we have provided links for both the Japanese and English versions of the materials. The Japanese version was utilized by the participants, who are Japanese EFL learners, while the English version is intended for readers of this manuscript who may not understand Japanese.

3.4. Target words

For each test, a total of 80 target items were chosen. Participants' understanding of these words was assumed to represent their phonological vocabulary proficiency relevant to real-life L2 listening experience. These target words were selected in the following procedure:

1. A speech corpus was developed from scripts of a retired version of the TOEIC Listening test. We opted for this test/material because the TOEIC test encapsulates various forms of L2

discourse, such as short sentences, conversations and monologues. A total of 2,731 tokens were identified.

2. Among them, the top 80 words that posed the most phonological challenges for Japanese EFL learners were selected as we eliminated words below 1,000 word families based on their word frequency in the BNC/COCA word family lists (Nation, 2012) and loadwords which could potentially facilitate L2 understanding (Uchihara et al., 2022).
3. We prioritized words that Japanese learners of English might find tricky, such as iambic words with numerous syllables, challenging segmentals like English [r] and [l], and consonant clusters (Saito, 2014).

The 80 words selected for the study comprised varied frequency profiles: the most common ones were within the top 2,000 word families, while the least frequent extended up to the top 8,000. As research suggests, this frequency span (i.e., 2,000 and 8,000) encompasses roughly 98% of words used in oral discourse (Nation, 2006) and is crucial for proficient L2 listening comprehension (Van Zeeland & Schmitt, 2013). Among these 80 words, a larger proportion were high-frequency (22 words within the 2,000 range and 35 in the 3,000 range). In contrast, fewer words were mid-frequency, with 13 in the 4,000 range and 10 spanning the 5,000–8,000 range. This distribution reflects the widely accepted notion that mastery of high-frequency vocabulary significantly influences L2 listening test outcomes. Such knowledge alone can explain more than half of the variability in overall L2 listening performance, especially among EFL listeners with limited international exposure, which characterizes the primary demographic of our study (Matthews, 2018). To minimize participants' explicit attention to target words, they first took the LJT (10–15 minutes), aural version of MR (5 minutes), and written version of MR (5 minutes) in this order.

3.4.1. *Automatized vocabulary test: LJT*

Building on the methodological paradigm in L2 grammar studies (Plonsky et al., 2020), the lexicosemantic judgment task (LJT) was developed in prior research (Saito et al., 2023; Uchihara et al., 2024). In this task, participants listened to 160 short sentences spoken by a female who speaks General American English. After hearing each one, they were asked to decide if the sentence was “semantically appropriate” or “semantically inappropriate” based on a single word in the sentence. To make sure participants listened to the whole sentence, the target word was not located at the beginning of the sentence. Sentences were kept simple, 4–8 words long, and always grammatically correct. Most of the words in these sentences (93%) came from the most common 1,000 words in English. Half of the sentences used the target word in a way that made sense, while the other half used it in a way that did not. For example, with the word “*estate*,” participants heard a sentence like “*My grandfather bought an estate*” (semantically appropriate), and “*My friend's estate was very kind*” (semantically inappropriate), which does not. All other parts of the sentence were clear and easy to understand. After the sentences were drafted, three English experts checked them to make sure they sounded semantically appropriate or inappropriate in the ways we intended. The chosen 160 sentences were played to participants in a random order. They got 1 point for each sentence they correctly judged, with the highest possible score being 160 points.

3.4.2. *Declarative vocabulary test: MR*

To our knowledge, empirical studies have yet to define the declarative dimension of L2 phonological vocabulary knowledge

comprehensively. Drawing on literature in skill acquisition theory (DeKeyser, 2017; Suzuki, 2023), L2 vocabulary assessment (Du et al., 2022), and word recognition (Coltheart & Rastle, 1994), we aim to define, justify, and methodologically approach this dimension within the context of the specific focus of the current investigation – i.e., adult EFL learners acquiring L2 phonological vocabulary toward successful L2 listening comprehension skills.

Skill acquisition theory in instructed SLA posits the declarative aspect of lexical knowledge primarily as form-meaning mapping, essential for acquisition and consolidation through practice (DeKeyser, 2017; Suzuki, 2023). To further elaborate this definition, according to Nation (2013), form-meaning mapping involves recognizing a word's sound and appearance. Although learners' behaviors could vary when their word knowledge is tested via written and aural modalities (McLean et al., 2015), Du et al. (2022) found that such modality effects could be mediated by L2 learners' proficiency levels. Under EFL classroom conditions (the main focus of the current study), learners' written performance generally precedes their aural recognition, especially from low-to-mid proficiency levels. Given the lack of phonetic training in many EFL classrooms all over the world, learners typically engage in form-meaning mapping without such foundational phonetic and phonological knowledge, leading to reliance on orthographic cues before transitioning to aural comprehension (Saito, 2021). This context-specific trajectory suggests three stages in developing L2 phonological vocabulary knowledge especially in EFL classroom settings:

- Stage 1: Non-recognition of L2 words (i.e., no knowledge).
- Stage 2: Recognition of word meanings in written form only.
- Stage 3: Comprehension of word meanings in both written and aural forms.

The three-stage model aligns with the dual route processing account of L2 word recognition (Coltheart & Rastle, 1994). According to this view, learners have both orthographic and phonological processing routes for retrieving words from their mental lexicon. Thus, when hearing spoken words, learners with underdeveloped aural lexicons may utilize their written vocabulary knowledge and phoneme–grapheme conversion rules to identify the heard words (i.e., spoken input → sound-to-spelling encoding → accessing written lexicons → word recognition). As learners' L2 proficiency and experience increase, they develop and directly rely on aural representations without accessing written representations (Milton & Masrai, 2021).

To accurately capture Japanese EFL learners' experiences with form-meaning mapping, we define declarative phonological vocabulary knowledge as the ability to link written and auditory word forms with their meanings. Following the developmental trajectory suggested by Du et al. (2022), we employed both written and aural formats in the MR task to provide a comprehensive evaluation of learners' phonological vocabulary knowledge, with and without orthographic cues. This methodological approach aligns with previous studies on vocabulary knowledge assessment and reflects the realistic learning processes encountered in EFL settings – initial learning without phonetic training, gradual familiarization with phonological aspects through practice, and eventual word recognition without orthographic assistance (Milton & Hopkins, 2006).

We are aware that some researchers advocate for exclusive use of aural tests as the most *cost-effective* method of assessing direct applicability to L2 listening comprehension (Cheng & Matthews,

2018). However, we emphasize that our study's primary goal is to explore the *developmental* aspect of L2 phonological vocabulary knowledge in EFL classrooms. An aural-only approach might not accurately represent learners in the interlanguage stage who understand words with orthographic cues but not without them (stages 2 but not 3). By incorporating both aural and written formats, we aim to reflect the comprehensive acquisitional process of L2 phonological vocabulary development in EFL classroom, recognizing the importance of capturing learners' transitional stages towards autonomous aural comprehension (i.e., learning L2 vocabulary with → without orthographic cues [Stages 2 →]).

For the aural version, the 80 target words were recorded from a female native speaker of American English. After each word, participants were asked to select its correct meaning from among four options (one correct answer and three distractors). All the choices belonged to the same grammatical category. Adapting the method used by McLean et al. (2015), each answer and distractor were translated into Japanese. This was done to reduce potential confusion and enhance comprehension for the Japanese participants. Three Japanese experts with extensive EFL teaching backgrounds assessed the multiple-choice options. Based on their feedback, necessary adjustments were made, especially concerning translation discrepancies in the answers and distractors. The aural prompts were presented in a randomized sequence. Upon completion of the aural task, participants moved on to the written format. Here, the same target words were displayed in randomized order, accompanied by their corresponding spellings. The multiple-choice options remained consistent with the aural version. This written version was included to check if participants truly understood the meanings of the target words or if they had merely struggled with auditory comprehension.

Each correct response earned one point, leading to a potential maximum of 160 points – 80 for both the aural and written sections.

3.5. Auditory processing measures

To assess participants' individual differences in processing both spectral and temporal dimensions of sound characteristics, we utilized two widely adopted auditory processing tasks: formant discrimination and amplitude risetime discrimination (Saito & Tierney, 2023). Within the sensory framework of language learning, it is posited that individuals vary in their perceptual abilities to encode fundamental acoustic characteristics – i.e., spectral and temporal information; and that this individual variation influences the rate and ultimate attainment of language learning. Spectral and temporal processing is instrumental to effective perception and learning of individual sounds (e.g., distinguishing duration and formant differences in vowel acquisition) and prosody (e.g., recognizing amplitude, duration, and pitch for stress and intonation). Such phonological proficiency is crucial for the recognition, learning, and automatization of lexicogrammar (for a comprehensive review, see Saito, 2023). Therefore, the formant discrimination and amplitude risetime discrimination tasks were selected to specifically measure participants' abilities to process spectral and temporal auditory information.

3.5.1. Stimuli

For the formant subtest, 101 complex nonspeech tones were created: one main stimulus (Level 0) and 100 comparison stimuli (Levels 1–100). Each had a length of 500 ms. To ensure smooth

sound transitions, 5-ms amplitude ramps were added at both the start and finish of every tone. The base frequency was fixed at 100 Hz, including harmonics reaching up to 3,000 Hz. Three formants were generated at 500 Hz, 1,500 Hz, and 2,500 Hz, using a standard formant filter method (Smith, 2007). The main tone had its second formant (F2) at 1,500 Hz. In contrast, the comparison tones were set between 1,502 and 1,700 Hz, with minor increments of around 2 Hz. For the amplitude rise time subtests, 101 tones were created, each having four harmonics, with the base frequency (F0) fixed at 330 Hz. Both the beginning and end of each tone had a 5-ms linear ramp. The duration of the initial amplitude ramp, 15 ms for the standard tone, varied between 10 and 300 ms for the others.

3.5.2. Procedure

In each trial during both tasks, participants listened to three nonspeech sounds in sequence and identified which sound (first or third) was different from the others. While the second sound was always fixed as the standard stimulus (Level 0), the first and third sounds comprised the comparison stimuli (Levels 0–100). The three sounds were identical except for one acoustic dimension (either formant [F2 = 1,502–1,700 Hz] or timing of initial amplitude risetime [10–300 ms]). Using Levitt's (1971) adaptive procedure, the size of the differences in the target acoustic dimension varied depending on participants' performance (the difference became smaller if they got correct responses and larger if they got incorrect responses). The resulting scores were scored at Levels 1–100 and smaller values indicated finer discrimination abilities. Participants' composite auditory processing scores were calculated by standardizing and averaging their formant and amplitude risetime scores. The descriptive results for auditory processing scores (raw, standardized) were summarized in Table 1.

3.6. Working memory measures

In Baddeley's (2000) framework, working memory is described as having four components: phonological loop, central executive, visuospatial sketchpad, and episodic buffer. In the current investigation, we focus on the two main components directly relevant to L2 listening comprehension: the phonological loop, which relates to how much information can be retained, and the central executive, which handles active use of this data. Following the methodological standards in L2 research (Li, 2016), these components were assessed through forward and backward digit span tasks. Following a number of empirical studies which adopted the same methodological decision (e.g., Olsthoorn et al., 2014), we used forward and backward digit span tasks (participants undertook the tasks in this order).

During these tasks, participants tried to remember a series of numbers and had to either recall them in the given sequence (forward span) or in a reversed sequence (backward span). Their responses were recorded via a keyboard. The series ranged in length from three up to 11 numbers, and for each length, participants had two tries. Each number was displayed for 500 ms on the computer screen. A participant's score on each task was determined by the length of the longest sequence they recalled correctly on both tries. Participants' composite working memory scores were calculated by standardizing and averaging their forward and backward digit span scores.¹

In the current investigation, our aptitude framework focused on domain-generalizability as opposed to domain-specificity. Therefore, we chose to employ digit span tasks that processed

Table 1. Biographical information, and L2 learning outcomes of participants

	<i>M</i>	<i>SD</i>	Range	95% CI	
				Bottom	Upper
A. Biographical backgrounds					
Age (years)	25.2	3.8	19–36	24.5	26.0
Age of EFL learning	3.0	2.2	.5–10	2.6	3.5
Length of EFL learning (hours)	21.6	4.1	16–36	20.8	22.4
Listening activities (%)	69.7	18.0	18.7–100	66.1	73.2
Reading activities (%)	24.1	21.8	2.3–89.6	19.9	28.5
Study abroad (presence)	Yes (<i>n</i> = 134), No (<i>n</i> = 352)				
Study abroad (years)	.112	.459	0–4.1	.079	.164
B. Aptitude backgrounds					
Composite auditory processing ^a (<i>Z</i> values)	.000	.744	–1.41–3.07	–.0670	–.0682
Formant discrimination ^a	34.6	18.4	2.5–99.0	33.0	36.3
Amplitude risetime discrimination ^a	31.6	20.7	3.2–97.6	29.7	33.3
Composite working memory (<i>Z</i> values)	.000	.720	–2.00–.77	–.655	.653
Forward span	7.2	1.2	4–8	7.1	7.3
Backward span	6.6	1.4	3–8	6.4	6.7

Note: ^aSmaller auditory processing scores indicating more precise perceptual abilities.

perceptual information visually rather than auditorily. These tasks were specifically designed to assess individuals' abilities to rehearse, memorize, and internally process such information as part of their inner speech mechanism. Importantly, we deliberately refrained from using working memory tasks that directly involved auditory materials (e.g., auditory working memory tasks). Our decision was informed by the potential confounding effect of individuals' lower-order auditory processing abilities, which have also been identified as critical determinants in various aspects of L2 learning (see Saito, 2023). To ensure a more accurate assessment of cognitive working memory abilities while controlling for auditory processing, we opted for non-audio digit span tasks and conducted separate tasks to evaluate auditory processing capabilities.

3.7. EFL experience questionnaire

The EFL Experience Questionnaire, adapted from Saito and Hanzawa (2016), gathered data on participants' past EFL experiences. It asked when they began learning L2 English (age of onset) and the number of hours they dedicated to form-oriented English lessons per week during elementary (Grades 1–6), junior high (Grades 7–9), and senior high school (Grades 10–12). The figures included the hours of EFL classes at cram schools as well. From this data, we calculated the total form-oriented EFL education hours for each participant. Given that many EFL students make efforts to access communicatively authentic input and engage in meaning-oriented EFL activities, participants were queried about the details of the participants' current EFL learning experience beyond form-oriented classes at school. Using a sliding scale, they indicated a percentage from 0% (never utilizing this approach) to 100% (always utilizing this approach) to describe their engagement in two key extracurricular learning activities at the time of the project. They featured (1) listening to aural materials (e.g., movies, YouTube, songs; Peters et al., 2019) and (2)

reading written materials (e.g., textbooks, novels, newspapers; De Wilde et al., 2021).

Lastly, participants were asked to report whether they had any experience studying abroad in English-speaking countries (e.g., USA, UK, Canada, Australia), and if so, for how long. To ensure that these experiences were substantive in terms of language learning opportunities, we only included stays abroad that lasted at least one month, excluding brief visits for family, friends, school, or solo travel without significant language learning opportunities. In the present study, the majority of participants had study-abroad experiences lasting less than one year (typical of study-abroad programs at universities in Japan). Consequently, we categorized the participants into two groups: those with study-abroad experience (*n* = 134) and those without such experience (*n* = 352). Participant demographics and experience profiles were summarized in Table 1.

3.8. Analyses

To address RQ1, we conducted comparative analyses to determine the extent of variance in participants' vocabulary accuracy and fluency scores as measured by two distinct test formats: the declarative (form-meaning mapping) and automatized test formats (lexicosemantic judgement). For RQ2, we employed a series of mixed-effects regression analyses to explore the relationships between participants' vocabulary scores and their individual differences in experience, aptitude, and metacognition.

4. Results

4.1. Declarative and automatized vocabulary performance

As presented in Table 2 and illustrated in Figure 1, the participants' phonological vocabulary performance was assessed both at the declarative (MR) and automatized levels (LJT). Both task

Table 2. Descriptive results of L2 phonological vocabulary knowledge

	M	SD	Range	95% CI	
				Bottom	Upper
A. Declarative knowledge					
MR ^a	131.1	17.4	58–160	129.6	132.7
B. Automatized knowledge					
LJT ^a	108.5	16.1	68–152	107.0	109.9

Notes: ^aA total of 160 points, with higher scores indicating more robust vocabulary knowledge.

^bSmaller values indicate less variation and greater stability in performance.

formats demonstrated adequate internal consistency: MR ($\alpha = .94$, [.93, .95]) and LJT ($\alpha = .93$, [.91, .95]). The descriptive results suggested that while many participants recognized the form-meaning aspects of the target words when reading and listening (evidenced by positive skewness), their ability to access this knowledge spontaneously was somewhat more constrained (evidenced by negative skewness). The results from the Kolmogorov–Smirnov tests confirmed that all the vocabulary scores (LJT, MR) diverged significantly from a normal distribution ($D = .082$ and $.167$, $p < .05$). In line with Larson-Hall’s (2015) field-specific guidelines, the scores from MR and LJT (with a mild skewness) underwent a log10 transformation. The directionality of all the transformed vocabulary scores follows the same trajectory, with higher scores indicating more advanced performance. Strong correlations were found between participants’ MR and LJT scores ($r = .601$, $p < .001$, [.540, .656]).

4.2. Relations between experience and aptitude

Participants’ profiles were coded based on several individual difference factors: quantity of experience (age of learning and length

of learning), quality of experience (listening and reading activities), and aptitude (auditory processing and working memory). To explore the inter-relationships among these factors, a series of Spearman correlations were performed. The alpha was set and adjusted to $p < .007$ to account for Bonferroni corrections across the seven comparisons. As displayed in Table 3, there were weak-to-moderate significant correlations among similarly themed variables. Specifically, age and length of learning correlated ($r = -.206$), suggesting that those who began learning earlier typically spent more time studying L2 English in classroom. There were correlations between conversation, listening, reading, and study-abroad activities, indicating that certain students are more involved in extracurricular activities. Interestingly, albeit weak, there were notable correlations between experience and aptitude. Participants who began learning at an earlier age exhibited more refined auditory processing ($r = .131$). Since none of the experience and aptitude variables exhibited strong correlations that might lead to multicollinearity issues, all were retained as predictors for L2 phonological vocabulary learning outcomes in subsequent analyses without any further data reduction.

4.3. Roles of experience and aptitude in phonological vocabulary development

The final objective of the statistical analyses concerns how participants’ phonological vocabulary development in EFL classrooms can be related to their prior and current learning backgrounds and aptitude profiles. To this end, a set of linear mixed effects regression analyses were performed on the two different dimensions of phonological vocabulary knowledge (declarative and automatized) in relation to a total of seven predictors via the lme4 package (Bates et al., 2021) in the R statistical environment (Version 4.3.1; R Core Team, 2023). The model incorporated five experience-related predictors: two reflecting the quantity of

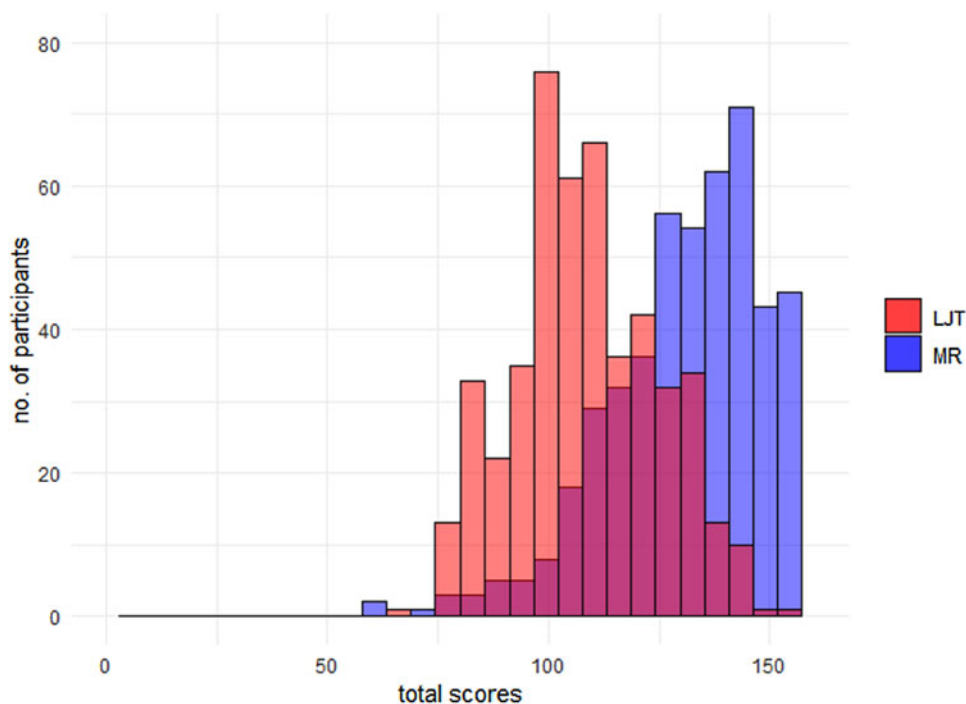


Figure 1. Graphical representation of participants’ vocabulary performance. Accuracy was notably higher in MR compared to LJT.

Table 3. Non-parametric correlations between experience and aptitude factors

	1	2	3	4	5	6
1. Age of learning						
2. Length of learning	-.206					
	<.001*					
3. Listening	-.019	.088				
	.695	.065				
4. Reading	-.090	.153	.240			
	.059	.001*	<.001*			
5. Study-abroad	.071	-.126	-.143	-.105		
	.141	.008	.003*	.028		
6. Auditory processing	.131	-.016	-.007	.058	.031	
	.007*	.731	.891	.228	.526	
7. Working memory	.053	-.014	-.074	-.014	.025	-.012
	.274	.757	.120	.770	.608	.794

Note: * for $p < .007$ (Bonferroni corrected).

form-oriented EFL experience (i.e., age and duration of EFL exposure) and three representing the quality of extracurricular EFL experience (i.e., listening and reading activities, along with study abroad experiences). Additionally, two aptitude-related predictors were included, targeting perception (i.e., auditory processing) and cognition (i.e., working memory). Individual learner (participants' ID) was included as a random factor.

- MODEL: $DV \sim \text{experience_factors} \times \text{test_type} + \text{aptitude_factors} \times \text{test_type} + (1|ID)$ (Table 4)

Significant main effects were observed for test type (MR vs. LJT; $p = .001$), reading activities ($p < .001$), study abroad experiences ($p = .023$), auditory processing ($p = .039$), and working memory ($p = .002$). Notably, interaction effects emerged as significant for age of onset of learning ($p < .001$), length of learning ($p = .035$), listening activities ($p = .008$), study abroad experiences ($p = .029$), and working memory ($p = .016$). To further disentangle these significant interaction effects, post-hoc analyses were conducted. These analyses aimed to discern how the continuous predictors (age of onset of learning, duration of

Table 4. Summary of mixed effects modeling analyses of experience- and aptitude-related factors and phonological vocabulary

Fixed effects	Estimate (b)	SE	t	p	Random effects	Variances	SD	$R^2_{\text{conditional}}$	R^2_{marginal}
Intercept	2.094	.015	137.160	<.001	Participant ID	.002	.047	.755	.369
Age of learning	<.001	.001	.426	.670	Residual	.001	.037		
Test type	-.040	.013	-3.007	.003*					
Length of learning	<.001	<.001	1.690	.092					
Listening activities	<.001	<.001	-.829	.407					
Reading activities	<.001	<.001	3.491	.001*					
Study-abroad	-.015	.007	-2.192	.029					
Auditory processing	-.007	.004	-1.698	.090					
Working memory	.012	.004	2.844	.005					
Age of learning: Test	-.003	.001	-3.840	<.001					
Length of learning: Test	<.001	<.001	-2.355	.019*					
Listening: Test	<.001	<.001	2.858	.004*					
Reading: Test	<.001	<.001	.268	.789					
Study-abroad: Test	-.014	.006	-2.278	.023*					
Auditory processing: Test	-.003	.004	-.925	.356					
Working memory: Test	-.008	.004	-2.169	.031*					

Note: * for $p < .05$.

Table 5. Post-hoc analyses summary of predictors for L2 phonological vocabulary

A. Continuous Predictors					
Predictor variables	Dependent variables	<i>r</i>	<i>p</i>	95% CI	
				Lower	Upper
Age of learning	MR	-.034	.474	-.127	.059
	LJT	-.180	<.001*	-.269	-.088
Length of learning	MR	.114	.017*	.021	.205
	LJT	.069	.145	-.024	.161
Listening	MR	.022	.635	-.070	.115
	LJT	.155	<.001*	.063	.245
Working memory	MR	.143	.001*	.054	.231
	LJT	.028	.535	-.061	.118
B. Categorical Predictors					
Group variables	Dependent variables	<i>t</i>	<i>p</i>	95% CI (mean differences)	
				Lower	Upper
Study-abroad (yes, no)	MR	2.461	.014*	.003	.028
	LJT	4.904	<.001*	.021	.049

Note: * for $p < .025$ (Bonferroni corrected).

learning, listening activities, and working memory) and the categorical predictor (study abroad) might differentially influence vocabulary scores across two different test conditions – MR and LJT.

As summarized in Table 5 and visually plotted in Figure 2, Pearson's product-moment correlation analyses revealed that the continuous predictors were differentially correlated with participants' MR and LJT scores (an alpha set to $p < .025$; Bonferroni corrections). Specifically, MR performance was significantly related to length of learning ($p = .017$) and working memory ($p = .001$), whereas LJT performance displayed significant correlations with age of learning ($p < .001$) and listening activities ($p < .001$). Independent *t*-tests were conducted to compare the vocabulary performance of those with and without study-abroad experience using the MR and LJT tests. The results (see Table 5 and Figure 2) indicated that participants with study-abroad experience significantly outperformed those without such experience in both tests ($p = .014$ for MR and $p < .001$ for LJT). However, according to the analyses of Cohen's *d*, the effect size of the difference in performance between the two groups was almost twice as large for LJT ($d = .545$) compared to MR ($d = .255$).

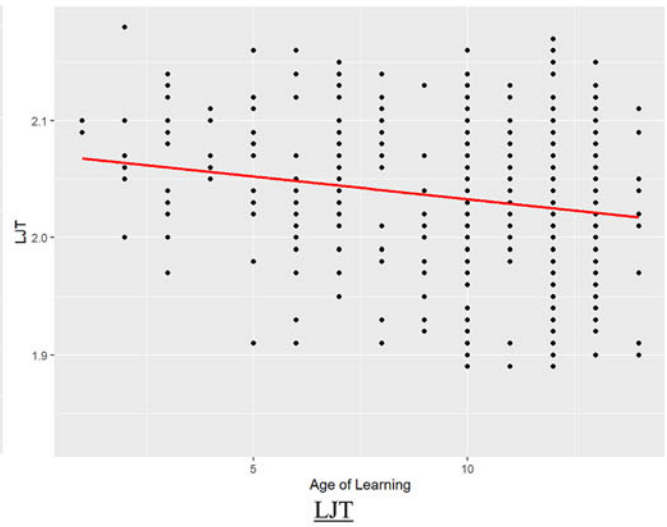
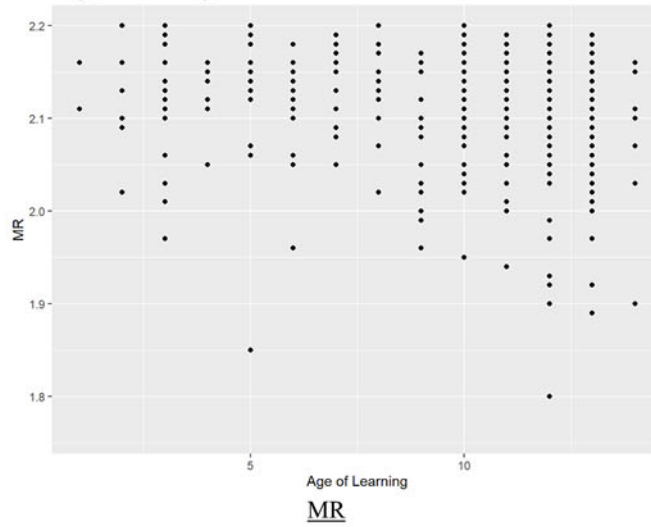
Finally, as study-abroad was considered a categorical variable, it is important to explore whether the duration of study-abroad influenced participants' declarative and automatized vocabulary knowledge. As a follow-up analysis, a mixed effects model was employed, incorporating participants' length of study abroad ($M = .12$ years, $SD = .45$ years, $Range = 0-4.1$ years). The analysis revealed significant interaction effects (but not main effects) of study abroad duration ($b = .001$, $SE = .005$, $t = 2.648$, $p = .008$). Correlation analyses indicated that the duration of study abroad significantly affected the LJT task ($r = .251$, $p < .001$) and showed a marginally significant impact on the MR task ($r = .105$, $p = .026$). All the data and R scripts presented here were summarized in Supporting Information-S2.

5. Discussion and future directions

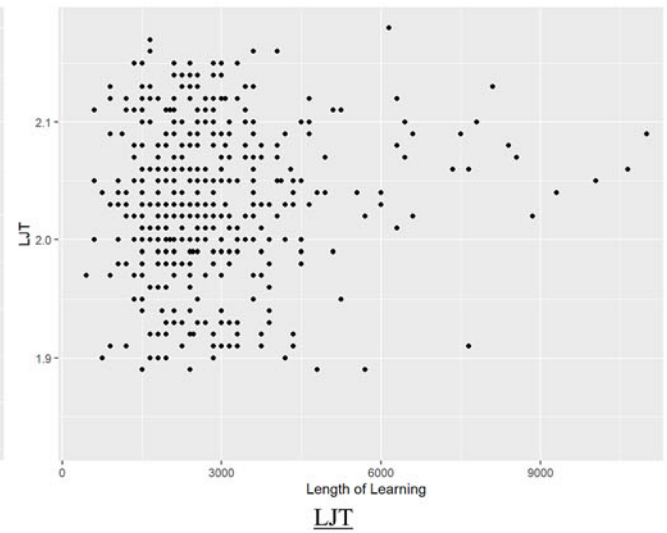
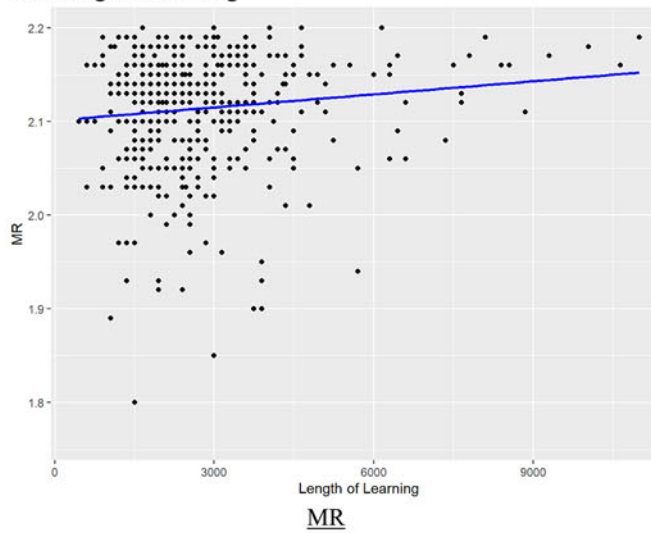
The present study investigated the following research questions: (1) what characterizes a total of 486 adult EFL learners' L2 phonological vocabulary knowledge – a key predictor of successful L2 listening comprehension; and (2) the interplay of experience and aptitude factors in shaping these outcomes. Unlike prior research (e.g., Muñoz, 2014), this study assessed L2 phonological vocabulary using two tasks: MR and LJTs. The MR task aimed to measure the declarative aspects of L2 phonological knowledge, reflecting an explicit and controlled processing of word form and meaning (McLean et al., 2015). In contrast, the LJT task gauged the automatized facets of this knowledge, indicating spontaneous and consistent access to target words within specific contexts (Saito et al., 2023; Uchihara et al., 2024). According to the skill acquisition theory for Instructed SLA (DeKeyser, 2017; Suzuki, 2023), L2 learners initially grasp the form and meanings of new words (declarative stage) and, through repeated practice, further refine this knowledge, allowing rapid and stable access based on the word's relationship with surrounding words (automatization stage; N. Ellis, 2006). This dichotomy between declarative and automatized knowledge reflects Nation's (2013) influential framework of word knowledge, which includes form-meaning mapping (understanding the sound, appearance, and significance of words) and context-in-use (insight into the circumstances of word usage).

With respect to RQ1 (i.e., the analyses of participants' L2 phonological vocabulary knowledge), findings indicated that participants demonstrated higher vocabulary performance when tested in MR than in LJT, mirroring previous studies (McLean et al., 2015; Uchihara et al., 2024). Drawing from the skill acquisition theory, these outcomes imply that while adult L2 learners might possess ample explicit vocabulary knowledge, they could face challenges in accessing it swiftly, stably, and automatically.

A-1. Age of learning



A-2. Length of learning



A-3. Listening activities

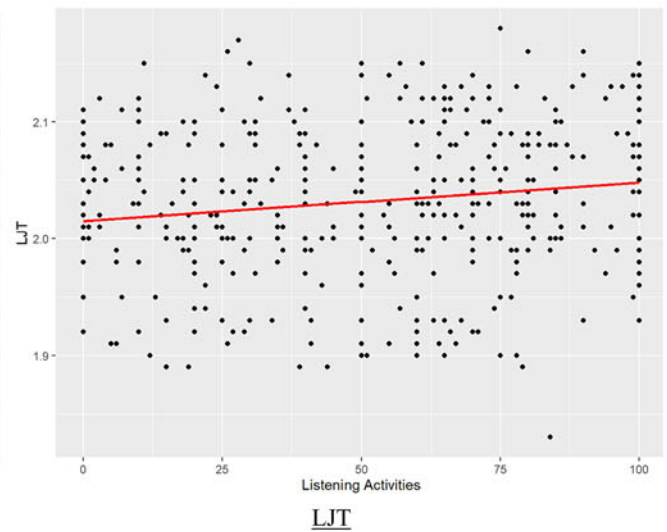
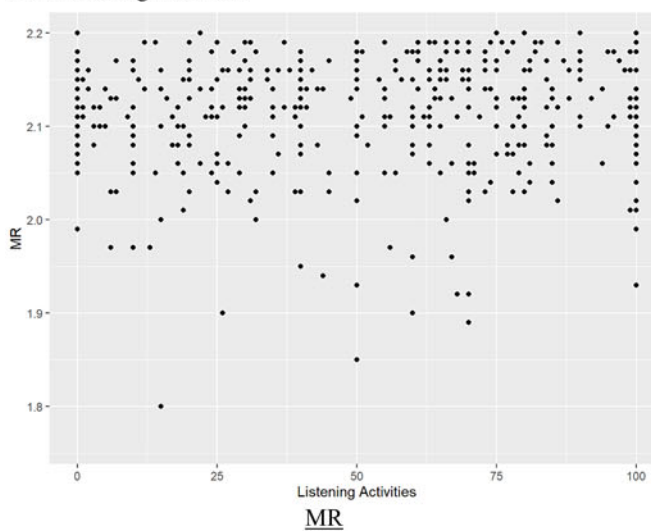
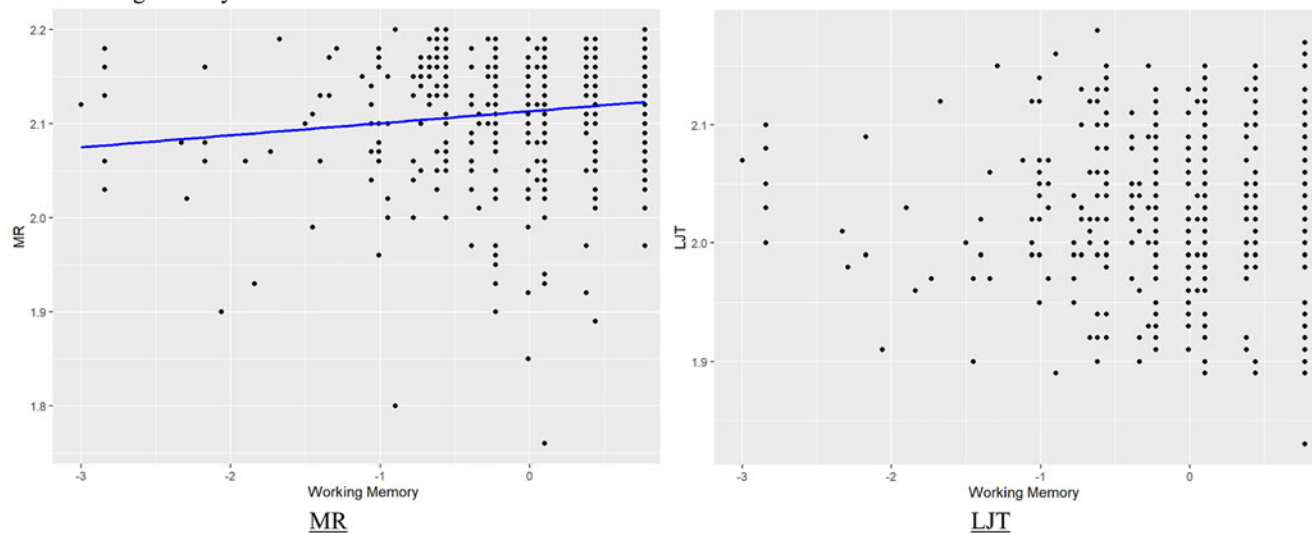


Figure 2. Graphical depiction of the relationship between experience and aptitude variables in relation to L2 phonological vocabulary performance in MR and LJT. Trendlines have been included to illustrate statistically significant links ($p < .05$) between vocabulary performance and the predictor variables. MR showed significant correlations with length of learning and working memory. In contrast, LJT demonstrated more evident effects from age of learning, listening activities, and study-abroad experiences.

A-4. Working memory



B. Study-abroad experiences

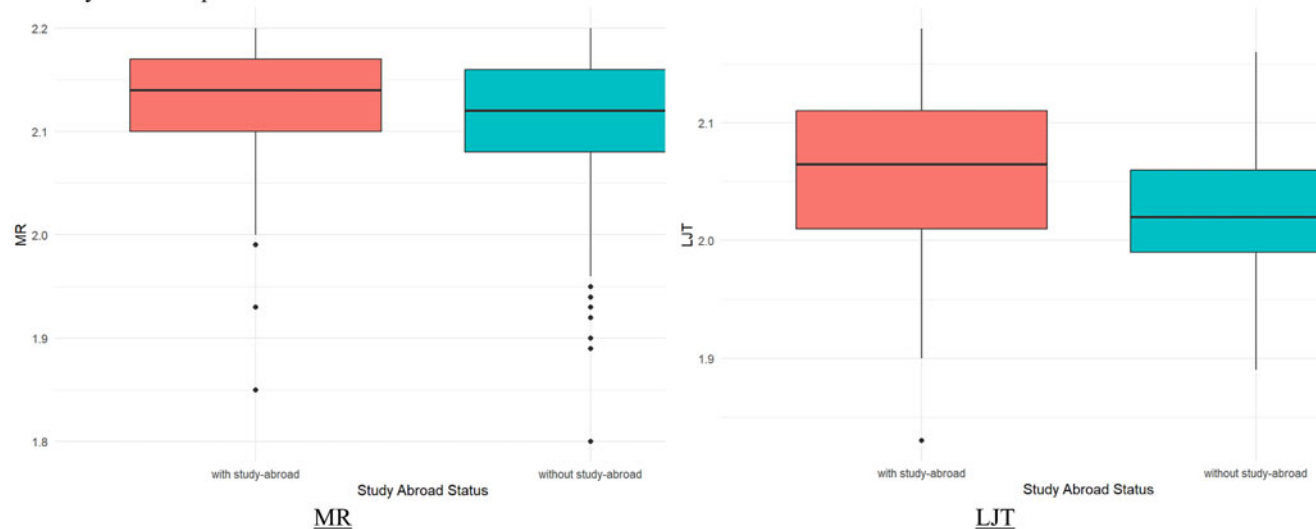


Figure 2. Continued.

Noteworthy is the differential correlation between MR and LJT scores with global L2 listening comprehension proficiency: LJT tends to hold a more potent predictive strength ($r = .6-.7$) than MR ($r = .4-.5$; Saito et al., 2023). Given past research tendencies to predominantly use MR formats in assessing L2 phonological vocabulary knowledge (Schmitt, 2019), there is a potential oversight in capturing learners' spontaneous, automatized knowledge – essential for real-world listening experiences. For effective L2 learning, it is imperative that teachers not only emphasize explicit word comprehension but also provide abundant practice to foster knowledge automatization.

With respect to RQ2 (i.e., the relationship between experience profiles, aptitude factors, and L2 phonological vocabulary development), the findings of mixed-effects regression analyses demonstrated three overall patterns. First and foremost, the development of L2 phonological vocabulary knowledge was equally driven by two experience variables – the amount of time that participants practiced an L2 outside form-oriented EFL lessons (especially through extensive reading activities; De Wilde et al.,

2020) and the presence of study-abroad experiences – and two aptitude variables: working memory (Elgort et al., 2020) and auditory processing (Saito et al., 2022). More importantly, two dimensions of L2 phonological vocabulary knowledge development were uniquely related to slightly different sets of experience and aptitude variables. Whereas those with more EFL training experience and greater working memory appeared to attain higher declarative L2 phonological vocabulary knowledge (measured via MR), those with an earlier age of learning, greater practice outside form-oriented lessons (especially through extensive listening activities), and study-abroad experiences tended to achieve greater automatized L2 phonological vocabulary knowledge.

To date, L2 phonological vocabulary knowledge has primarily been assessed using various declarative measures, such as MR (McLean et al., 2015) and meaning recall (Cheng et al., 2022). While some studies have explored the effects of intentional training (e.g., form-meaning mapping; Uchihara et al., 2022) and incidental exposure (e.g., TV watching; Peters & Webb, 2018), few have delved into the role of instruction in the development of

automatized phonological vocabulary knowledge (Schmitt, 2019; but see DeKeyser, 1997 for long-term effects of training on the acquisition of *morphosyntactic* features in a miniature language). Suzuki and Elgort's (2023) comprehensive review of measurement practices for automatized L2 knowledge underscores the limited attention given to the creation and utilization of tasks assessing automaticity in auditory lexical processing. To our knowledge, this study is the first initiative to reveal the experiential and perceptual-cognitive foundations of automatized (in contrast to declarative) phonological vocabulary knowledge. Consistent with the skill acquisition theory – which posits instructed L2 acquisition as transitioning from declarative to procedural and automatized knowledge – our findings offer fresh insights about the influence of long-term EFL training, extracurricular activities, and aptitude in the extant literature.

On one hand, given that EFL instruction typically encompasses several hours of form-oriented lessons, the duration of such exposure may be tied to the growth of declarative knowledge (McLean et al., 2015). To further optimize L2 learning, many EFL learners pursue extracurricular practice, often through written modalities such as extensive reading (De Wilde et al., 2020). While these activities undoubtedly influence L2 vocabulary acquisition, incorporating more auditory materials (e.g., extensive listening and watching) might specifically expedite the development of automatized vocabulary knowledge (Peters et al., 2019). Since real-life comprehension and speaking opportunities are paramount for the automatization of L2 knowledge (DeKeyser, 2007), study-abroad experiences appear to be profoundly linked to the enhancement of automatized phonological vocabulary knowledge, perhaps even more so than declarative phonological vocabulary knowledge (Muñoz, 2014).

On the other hand, given the potential limitations in the quantity and quality of EFL experience in classroom environments compared to naturalistic and immersion settings (Larson-Hall, 2008), learners with strong aptitude profiles might derive more benefits. These individuals can optimize each learning opportunity, leading to more substantial gains even with limited exposure to the target language (Wen & Skehan, 2021). Working memory appears to play a pivotal role in the development of accurate and fluent access to declarative vocabulary knowledge (Elgort et al., 2018; Ruiz et al., 2021). Yet, its role might be less obvious during the later stages of L2 vocabulary acquisition, such as use-in-context and automatization (Nation, 2013). It is important to note that in our current datasets, while we observed significant main effects for auditory processing and working memory, we did not pinpoint significant aptitude predictors specifically tied to automatized L2 phonological vocabulary knowledge. This underscores two potential interpretations concerning aptitude and automatization in L2 phonological vocabulary learning. First, the processes of automatization might be more influenced by the volume of repetitive, meaning-focused practice than inherent aptitude factors (Suzuki, 2021; cf. Ruiz et al. for the lack of working memory effects in meaning-oriented instruction). Alternatively, the development of automatized L2 knowledge might align with different aptitude dimensions, such as implicit statistical learning (Linck et al., 2013).

The current study revealed that participants who began learning at an earlier age exhibited more advanced automatized (as opposed to declarative) phonological vocabulary knowledge. Considering the extensive scholarly debates regarding the influence of age on L2 learning within classroom environments (e.g., Larson-Hall, 2008 vs. Muñoz, 2014), these findings seem

unexpected and should be approached with caution. Parsing the effects of age on EFL learning outcomes is challenging, mainly because the age of onset for learning often correlates and overlaps with the duration of learning (those who start earlier often spend more time engaged in EFL education). However, it is crucial to remember that the results of the current study found that participants who began their L2 journey earlier typically demonstrated superior auditory processing. As other research has indicated, auditory processing is often linked to age-related factors (e.g., chronological age, age of L2 acquisition; Skoe et al., 2015). Thus, it is plausible to suggest that early starters, leveraging their refined and adaptable auditory processing capabilities, can optimize their EFL experiences, leading to more automatic execution of vocabulary knowledge. See also Elgort and Warren (2014) for the significant effects of age of learning on EFL learners' tacit vocabulary knowledge (measured via lexical decision task) but not on explicit knowledge (measured via meaning generation task).

In the present study, we conducted an initial exploration into the mechanisms underlying the development of both declarative and automatized phonological vocabulary knowledge in EFL classrooms, paving the way for several future research avenues.

- Although the current study indicates that declarative and automatized phonological vocabulary knowledge can be two different learning phenomena related to different affecting variables, our research was cross-sectional. Future studies might benefit from a longitudinal design to validate our findings and understand how different types of training (e.g., intentional vs. incidental) might impact the development of both declarative and automatized knowledge over time. In a related vein, we recently conducted a follow-up study (Saito et al., [forthcoming-a](#)) to examine the differential effects of two types of phonological vocabulary training – MR with feedback versus lexicosemantic judgement with feedback. This study aims to confirm how these training methods differentially impact global L2 listening proficiency (Saito et al., [forthcoming-a](#)). These findings could provide additional longitudinal evidence supporting the distinction between declarative and automatized phonological vocabulary, as well as the relevant assessment tasks.
- While our focus was on EFL settings, which can often be highly form-centric with limited opportunities for automatization, a potential future direction could involve studying long-term, advanced L2 learners in immersive settings, where the distinctions between declarative and automatized phonological vocabulary might be more clearly observed.
- Although numerous predictors of L2 phonological vocabulary knowledge showed statistical significance, their interpretation requires caution. The correlation coefficients for these predictors ranged between .1 and .2. For instance, the effect of working memory on MR was $r = .143$, which is notable but comparatively small, especially when considering that the average effect of aptitude in L2 learning is typically small-to-medium ($r = .3-.4$; as reported by Li, 2016).
- A reviewer suggested the inclusion of a time limit as an additional methodological variable to potentially enhance the validity of the lexicosemantic judgement task in measuring automatized vocabulary knowledge. While imposing a time constraint on each stimulus might predominantly reflect increased processing speed rather than the stabilization aspect of automatization (Segalowitz, 2010), we have conducted a follow-up study. In this study, we aimed to establish an appropriate time limit for contextually appropriate versus

inappropriate sentences. This was based on the performance of native speakers and L2 speakers with varying proficiency levels. The detailed results of this investigation will be reported in a separate publication (Saito et al., [forthcoming-b](#)). Additionally, the timed LJT will be made available in L2 Speech Tools (Mora-Plaza et al., 2022).

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S1366728924000609>.

Data availability statement. The test materials used in the current study are deposited in L2 Speech Tools (<http://doi.org/10.17616/R31NJNAX>).

Acknowledgements. We acknowledge the invaluable assistance of the following team members in data collection and analysis: Kotaro Takizawa, Yui Suzukida, Satsuki Kurosawa, Masaki Eguchi, Noriaki Mikajiri, Izumi Hosaka, Noriko Nakanishi, Nobuhiro Kamiya, Konstantinos Macmillan, and Magdalena Kachlicka. We also thank the three anonymous reviewers for their insightful feedback. This project was funded by the Leverhulme Trust Research Grant (RPG-2019-039), the UK-ISPF Research Grant (1185702223), and the UCL-TU Strategic Partner Fund.

Competing interests. The authors declare no conflicting interests.

Note

¹ To examine the influence of potentially outlier participants, the analyses were redone after removing those with *Z* scores of 2.0. However, the results remained consistent with our initial findings.

References

- Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, 59(2), 249–306.
- Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, 52, 388–407.
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2021). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bisson, M. J., Kukona, A., & Lengeris, A. (2021). An ear and eye for language: Mechanisms underlying second language word learning. *Bilingualism: Language and Cognition*, 24(3), 549–568.
- Carroll, J. B., & Sapon, S. M. (1959). *Modern language aptitude test*.
- Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, 35(1), 3–25. <https://doi.org/10.1177/0265532216676851>
- Cheng, J., Matthews, J., Lange, K., & McLean, S. (2022). Aural single-word and aural phrasal verb knowledge and their relationships to L2 listening comprehension. *TESOL Quarterly*. <https://doi.org/10.1002/tesq.3137>
- Coltheart, M., & Rastle, K. (1994). Serial processing in reading aloud: Evidence for dual-route models of reading. *Journal of Experimental Psychology: Human Perception and Performance*, 20(6), 1197.
- Dang, T. N. Y., Webb, S., & Coxhead, A. (2022). Evaluating lists of high-frequency words: Teachers' and learners' perspectives. *Language Teaching Research*, 26, 617–641. <https://doi.org/10.1177/1362168820911>
- DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition*, 19(2), 195–221.
- DeKeyser, R. (Ed.) (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge University Press.
- DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen, & M. Sato (Eds.), *The Routledge handbook of second language acquisition* (pp. 15–32). Routledge.
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2020). Learning English through out-of-school exposure. Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, 23(1), 171–185.
- De Wilde, V., Brysbaert, M., & Eyckmans, J. (2021). Young learners' L2 English after the onset of instruction: Longitudinal development of L2 proficiency and the role of individual differences. *Bilingualism: Language and Cognition*, 24(3), 439–453. <https://doi.org/10.1017/S1366728920000747>
- Du, G., Hasim, Z., & Chew, F. P. (2022). Contribution of English aural vocabulary size levels to L2 listening comprehension. *International Review of Applied Linguistics in Language Teaching*, 60(4), 937–956. <https://doi.org/10.1515/iral-2020-0004>
- Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, 64(2), 365–414. <https://doi.org/10.1111/lang.12052>
- Elgort, I., Candry, S., Boutorwick, T. J., Eyckmans, J., & Brysbaert, M. (2018). Contextual word learning with form-focused and meaning-focused elaboration. *Applied Linguistics*, 39(5), 646–667. <https://doi.org/10.1093/applin/amw029>
- Elgort, I., Beliaeva, N., & Boers, F. (2020). Contextual word learning in the first and second language: Definition placement and inference error effects on declarative and nondeclarative knowledge. *Studies in Second Language Acquisition*, 42(1), 7–32. <https://doi.org/10.1111/lang.12335>
- Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, 27(1), 1–24. <https://doi.org/10.1093/applin/ami038>
- Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42(3), 375–396.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27(2), 141–172. <https://doi.org/10.1017/S0272263105050096>
- Faretta-Stutenberg, M., & Morgan-Short, K. (2018). The interplay of individual differences and context of learning in behavioral and neurocognitive second language development. *Second Language Research*, 34(1), 67–101. <https://doi.org/10.1177/0267658316684903>
- González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *ITL-International Journal of Applied Linguistics*, 166(1), 94–126. <https://doi.org/10.1075/itl.166.1.03fer>
- Goswami, U. (2015). Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience*, 16(1), 43–54. <https://www.nature.com/articles/nrn3836>
- Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, 35(3), 423–449. <https://doi.org/10.1017/S0272263113000041>
- Hamada, Y., & Yanagawa, K. (2023). Aural vocabulary, orthographic vocabulary, and listening comprehension. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2022-0100>
- Hui, B., & Godfroid, A. (2021). Testing the role of processing speed and automaticity in second language listening. *Applied Psycholinguistics*, 42(5), 1089–1115. <https://doi.org/10.1017/S0142716420000193>
- JACET SLA (2013). *Second language acquisition and language teaching*. Kaitakusha.
- Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning*, 67(3), 631–664. <https://doi.org/10.1111/lang.12242>
- Kachlicka, M., Saito, K., & Tierney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representation of sound. *Brain and Language*, 192, 15–24. <https://doi.org/10.1016/j.bandl.2019.02.004>
- Kalashnikova, M., Goswami, U., & Burnham, D. (2019). Sensitivity to amplitude envelope rise time in infancy and vocabulary development at 3 years: A significant relationship. *Developmental Science*, 22(6), e12836. <https://doi.org/10.1111/desc.12836>
- Kamiya, N. (2022). The limited effects of visual and audio modalities on second language listening comprehension. *Language Teaching Research*. <https://doi.org/10.1177/136216882210962>

- Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research*, 24(1), 35–63. <https://doi.org/10.1177/02676583070829>
- Larson-Hall, J. (2015). *A guide to doing statistics in second language research using SPSS and R* (2nd ed.). Routledge.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477. <https://doi.org/10.1121/1.1912375>
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, 38(4), 801–842. <https://doi.org/10.1017/S027226311500042X>
- Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., & Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, 63, 530–566. <https://doi.org/10.1111/lang.12011/abstract>
- Lu, C., & Dang, T. N. Y. (2023). Effect of L2 exposure, length of study, and L2 proficiency on EFL learners' receptive knowledge of form–meaning connection and collocations of high-frequency words. *Language Teaching Research*. <https://doi.org/10.1177/13621688231155820>
- Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, 72, 23–36. <https://doi.org/10.1016/j.system.2017.10.005>
- McLean, S., Hogg, N., & Kramer, B. (2014). Estimations of Japanese university learners' English vocabulary sizes using the vocabulary size test. *Vocabulary Learning and Instruction*, 3(2), 47–55. <http://doi.org/10.7820/vli.v03.2.mclean.et.al>
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(6), 741–760. <https://doi.org/10.1177/1362168814567889>
- Meara, P. (2005). *Llama language aptitude tests: The manual*. Lognostics.
- Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *The Canadian Modern Language Review*, 63(1), 127–147. <https://doi.org/10.3138/cmlr.63.1.127>
- Milton, J., & Masrai, A. (2021). Vocabulary and listening. In J. Clenton, & P. Booth (Eds.), *Vocabulary and the four skills: Pedagogy, practice, and implications for teaching vocabulary* (pp. 45–59). Routledge.
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, 130(4), 621. <https://psycnet.apa.org/buy/2001-05320-003>
- Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). *Tools for second language speech research and teaching*. <http://sla-speech-tools.com>. <http://doi.org/10.17616/R31NJNAX>
- Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching*, 49, 113–133. <https://doi.org/10.1515/iral.2011.006>
- Muñoz, C. (2014). Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics*, 35(4), 463–482. <https://doi.org/10.1093/applin/amu024>
- Nation, I. (2006). *How large a vocabulary is needed for reading and listening?* *Canadian Modern Language Review*, 63(1), 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2012). *The BNC/COCA word family lists*. <https://www.wgtn.ac.nz/lals/about/staff/paul-nation>
- Nation, I. S. P. (2013). *Learning vocabulary in another language*. Cambridge University Press.
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. <https://doi.org/10.1037/0033-295X.115.2.357>
- O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition*, 29(4), 557–581. <https://doi.org/10.1017/S027226310707043X>
- Olsthoorn, N. M., Andringa, S., & Hulstijn, J. H. (2014). Visual and auditory digit–span performance in native and non–native speakers. *International Journal of Bilingualism*, 18(6), 663–673. <https://doi.org/10.1177/1367006912466314>
- Perez, M. M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 42(4), 749–773. <https://doi.org/10.1017/S0272263119000706>
- Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577. <https://doi.org/10.1017/S0272263117000407>
- Peters, E., Noreillie, A. S., Heylen, K., Bulté, B., & Desmet, P. (2019). The impact of instruction and out-of-school exposure to foreign language input on learners' vocabulary knowledge in two languages. *Language Learning*, 69(3), 747–782. <https://doi.org/10.1111/lang.12351>
- Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, 36(4), 583–621. <https://doi.org/10.1177/026765831982>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ruiz, S., Rebuschat, P., & Meurers, D. (2021). The effects of working memory and declarative memory on instructed second language vocabulary learning: Insights from intelligent CALL. *Language Teaching Research*, 25(4), 510–539. <https://doi.org/10.1177/13621688198728>
- Saito, K. (2014). Experienced teachers' perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics*, 24(2), 250–277. <https://doi.org/10.1111/ijal.12026>
- Saito, K. (2019). Individual differences in second language speech learning in classroom settings: Roles of awareness in the longitudinal development of Japanese learners' English/ɪ/pronunciation. *Second Language Research*, 35(2), 149–172. <https://doi.org/10.1177/02676583187683>
- Saito, K. (2021). What characterizes comprehensible and native-like pronunciation among English-as-a-second-language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly*, 55(3), 866–900. <https://doi.org/10.1002/tesq.3027>
- Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing auditory precision hypothesis-L2. *Language Teaching*, 56, 522–538. <https://doi.org/10.1017/S0261444822000453>
- Saito, K., & Hanzawa, K. (2016). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics*, 37(4), 813–840. <https://doi.org/10.1017/S0142716415000259>
- Saito, K., & Tierney, A. (2023). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*. <https://doi.org/10.1017/S027226312200047X>
- Saito, K., Sun, H., Kachlicka, M., Alayo, J. R. C., Nakata, T., & Tierney, A. (2022). Domain-general auditory processing explains multiple dimensions of L2 acquisition in adulthood. *Studies in Second Language Acquisition*, 44(1), 57–86. <https://doi.org/10.1017/S0272263120000467>
- Saito, K., Uchihara, T., Takizawa, K., & Suzukida, Y. (2023). Individual differences in L2 listening proficiency revisited: Roles of form, meaning, and use aspects of phonological vocabulary knowledge. *Studies in Second Language Acquisition*, 1–27. <https://doi.org/10.1017/S027226312300044X>
- Saito, K., Uchihara, T., Takizawa, K., & Suzukida, Y. (forthcoming-a). *Declarative and automatized phonological vocabulary knowledge: A training study*.
- Saito, K., Hosaka, I., Suzukida, Y., Takizawa, K., & Uchihara, T. (forthcoming-b). *Timed vs. untimed lexicosemantic judgement task for measuring automatized phonological vocabulary knowledge*.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. <https://doi.org/10.1177/13621688080899>
- Schmitt, N. (2019). Understanding vocabulary acquisition, instruction, and assessment: A research agenda. *Language Teaching*, 52(2), 261–274. <https://doi.org/10.1017/S0261444819000053>
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Skoe, E., Krizman, J., Anderson, S., & Kraus, N. (2015). Stability and plasticity of auditory brainstem function across the lifespan. *Cerebral Cortex*, 25(6), 1415–1426. <https://doi.org/10.1093/cercor/bht311>

- Smith, J. O. (2007). *Introduction to digital filters: with audio applications* (Vol. 2), [Published online]. W3k Publishing.
- Suzuki, Y. (2021). Individual differences in memory predict changes in breakdown and repair fluency but not speed fluency: A short-term fluency training intervention study. *Applied Psycholinguistics*, 42(4), 969–995. <https://doi.org/10.1017/S0142716421000187>
- Suzuki, Y. (Ed.). (2023). *Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology*. Routledge. <https://doi.org/10.4324/9781003414643>
- Suzuki, Y., & Elgort, I. (2023). Measuring automaticity in second language comprehension. In Y. Suzuki (Ed.), *Practice and automatization in second language research* (pp. 206–234). Routledge.
- Tagarelli, K., Mota, M., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. *Proceedings of the annual meeting of the cognitive science society*, 33 (No. 33). <https://escholarship.org/uc/item/0r55c3fk>
- Taguchi, N. (2011). Teaching pragmatics: Trends and issues. *Annual Review of Applied Linguistics*, 31, 289–310. <https://doi.org/10.1017/S0267190511000018>
- Uchihara, T., Webb, S., Saito, K., & Trofimovich, P. (2022). The effects of talker variability and frequency of exposure on the acquisition of spoken word knowledge. *Studies in Second Language Acquisition*, 44(2), 357–380. <https://doi.org/10.1017/S0272263121000218>
- Uchihara, T., Saito, K., Takizawa, K., Kurokawa, S., & Suzukida, Y. (2024). Declarative and automatized phonological vocabulary knowledge: Recognition, recall, lexicosemantic judgment, and listening-focused employability of second language words. *Language Learning*. <https://doi.org/10.1111/lang.12668>
- Vafae, P., & Suzuki, Y. (2020). The relative significance of syntactic knowledge and vocabulary knowledge in second language listening ability. *Studies in Second Language Acquisition*, 42(2), 383–410. <https://doi.org/10.1017/S0272263119000676>
- Vandergrift, L., & Baker, S. C. (2018). Learner variables important for success in L2 listening comprehension in French immersion classrooms. *The Canadian Modern Language Review*, 74(1), 79–100. <https://doi.org/10.3138/cmlr.3906>
- van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, 34(4), 457–479. <https://doi.org/10.1093/applin/ams074>
- Wallace, M. P. (2022). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*, 72(1), 5–44. <https://doi.org/10.1111/lang.12424>
- Webb, S. A., & Chang, A. C. S. (2012). Second language vocabulary growth. *RELC Journal*, 43(1), 113–126.
- Wen, Z. E., & Skehan, P. (2021). Stages of acquisition and the P/E model of working memory: Complementary or contrasting approaches to foreign language aptitude? *Annual Review of Applied Linguistics*, 41, 6–24. <https://doi.org/10.1017/S0267190521000015>
- Yashima, T., Zenuk-Nishide, L., & Shimizu, K. (2004). The influence of attitudes and affect on willingness to communicate and second language communication. *Language Learning*, 54(1), 119–152. <https://doi.org/10.1111/j.1467-9922.2004.00250.x>
- Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*. <https://doi.org/10.1177/1362168820913998>