Research Methods in Applied Linguistics 2 (2023) 100051

Contents lists available at ScienceDirect

Research Methods in Applied Linguistics

journal homepage: www.elsevier.com/locate/rmal

fMRI reveals the dynamic interface between explicit and implicit knowledge recruited during elicited imitation task

Yuichi Suzuki[a,†,*], Hyeonjeong Jeong[b,†], Haining Cui[c], Kiyo Okamoto[b], Ryuta Kawashima[b], Motoaki Sugiura[b]

[a] Kanagawa University, Japan
[b] Tohoku University, Japan
[c] McGill University, Canada

ARTICLE INFO

Keywords:
Explicit and implicit knowledge
Declarative and procedural memory
Elicited imitation task
Automaticity
fMRI

ABSTRACT

Development of valid tasks that tap into implicit knowledge is a prerequisite for understanding the interface between explicit and implicit grammatical knowledge in second language (L2) acquisition. However, the extent to which elicited imitation tasks (EITs) draw on implicit or/and explicit knowledge has been a subject of controversy, due in part to the limitations of behavioral methods. To overcome this drawback, in this study, we used functional magnetic resonance imaging (fMRI) to examine the neural circuits underlying explicit and implicit knowledge (i.e., declarative and procedural memory) during the listening and speaking phases of an EIT performed by advanced L2 speakers of Japanese living in Japan. While the behavioral data suggest that the EIT primarily draws on automatized (speeded-up) explicit knowledge, the neuroimaging data revealed learners' dynamic use of explicit and implicit knowledge during its comprehension and production phases. Higher explicit knowledge scores (derived from a metalinguistic knowledge task) were associated with greater declarative memory (left hippocampus) activation during the speaking EIT phase, indicating a prominent role of explicit knowledge in production. During the listening phase, however, higher explicit knowledge scores predicted lower activation in declarative memory (left hippocampus) and higher activation in procedural memory (left inferior frontal gyrus), suggesting that explicit knowledge plays both inhibitory and facilitative role in the use of implicit knowledge for comprehension. Taken together, these findings suggest that advanced L2 speakers utilize their explicit and implicit knowledge efficiently and dynamically—characterized as a hallmark of automaticity—for comprehension and production during the EIT.

Introduction

As elicited imitation tasks (EITs) require oral (re)production of a previously heard stimulus sentence, they are widely used for assessing second language (L2) proficiency. Despite their simplicity, such tasks require integration of listening (decoding auditory input for comprehension) and speaking (reconstructing the sentence from the semantic proposition to utter meaningful speech) skills, as well as recruitment of multiple aspects of linguistic knowledge, such as phonology, lexicon, and grammar. Moreover, findings

* Corresponding author at: Faculty of Cross-Cultural and Japanese Studies, Kanagawa University, 4-5-3 Minatomirai, Nishi-ku, Yokohama-shi, Kanagawa 220-8739, JAPAN.
E-mail address: szky819@kanagawa-u.ac.jp (Y. Suzuki).
† The first and second authors (YS and HJ) equally contributed to this work.

https://doi.org/10.1016/j.rmal.2023.100051
Received 9 January 2023; Received in revised form 5 April 2023; Accepted 6 April 2023

yielded by recent meta-analyses synthesizing several decades of research on EIT (e.g., Kostromitina & Plonsky, 2022; Yan et al., 2016) lend credence to EIT as a valid measure of L2 proficiency in the second language acquisition (SLA) context.

While EIT has been established as a useful instrument for assessing global L2 proficiency, it is also utilized in SLA research to assess grammatical knowledge (Erlam, 2006). In SLA research, grammatical knowledge is often distinguished based on awareness (e.g., DeKeyser, 2003; N. Ellis, 2015): explicit knowledge (knowledge that one is aware of) and implicit knowledge (knowledge that lacks conscious awareness). Because L2 researchers have vested interest in the interface of explicit and implicit knowledge and its practical implications (e.g., the extent to which explicit instruction can facilitate the acquisition of L2 knowledge that can be used for fluent communication), L2 researchers have started investigating the construct validity of EIT as a part of the ongoing collaborative effort to develop valid and reliable explicit and implicit knowledge tests (for recent reviews, see Isbell & Rogers, 2021; Roehr-Brackin, 2022).

Earlier validation studies have yielded evidence supporting the EIT use as a measure of implicit knowledge (e.g., Bowles, 2011; R. Ellis, 2009; Erlam, 2006; Zhang, 2015), which is challenged by more recent findings suggesting that explicit knowledge is primarily deployed during EITs (e.g., Spada, Shiu, & Tomita, 2015; Suzuki & DeKeyser, 2015). These discrepancies might be due to the nature of evidence provided by behavioral methods, as prior empirical approaches for examining EIT validity tended to rely on (a) retrospective questionnaires gauging respondents' levels of awareness of target structures embedded in the EIT (e.g., Erlam & Wei, 2021; Granena, 2016; Spada et al., 2015) and/or (b) comparing their performance on EIT and other tests such as time-pressured acceptability judgement task (e.g., R. Ellis, 2009) or real-time grammar comprehension task (e.g., Suzuki & DeKeyser, 2015).

Although such behavioral data is useful, it cannot be analyzed to directly infer the type of underlying linguistic knowledge. Therefore, new methods are required to directly demonstrate the dynamic use of explicit and implicit knowledge during EITs. Hence, as a part of this investigation, we subjected our participants to functional magnetic resonance imaging (fMRI) to examine the neural processes activated during EIT performance. Given that neurobiological models of memory postulate presence of brain circuits involved in declarative and procedural memory that presumably underlie explicit and implicit knowledge (Paradis, 2009; Ullman, 2020), neuroimaging L2 studies are increasingly being conducted to distinguish between explicit and implicit knowledge (Morgan-Short et al., 2015; Suzuki et al., 2022). To aid in this endeavor, we employed fMRI to scrutinize the neural and behavioral evidence and ascertain the extent to which participants rely on their explicit and implicit knowledge during an EIT.

Literature review

Awareness during EIT

EITs adopted to measure implicit knowledge have three unique methodological features (e.g., Erlam, 2006). First, ungrammatical as well as grammatical sentences are embedded in the EIT stimuli, enabling researchers to investigate test-takers' ability to spontaneously correct ungrammatical sentences. Second, after hearing each stimulus, learners are prompted to answer a comprehension question (e.g., plausibility judgement of the stimulus sentence) so that their attention is directed to meaning rather than linguistic form and/or sentence grammaticality. Third, time pressure is imposed during the EIT production phase, potentially limiting access to explicit knowledge.

As implicit knowledge is typically assessed using an awareness criterion, researchers often use retrospective questionnaires to determine the extent to which L2 speakers are aware of the linguistic structures embedded in EIT stimuli (e.g., Chrabaszcz & Jiang, 2014; Granena, 2016; Spada et al., 2015). For instance, in their study involving Chinese learners of English, Granena (2016) administered an EIT comprising grammatical and ungrammatical sentences featuring English plural-marking and subject–verb agreement. In order to probe their awareness of grammatical errors, participants completed a retrospective questionnaire immediately after the EIT. The results showed that almost all L2 speakers reported noticing (95%) and correcting (92%) the errors incorporated in the test. These findings suggest that EITs primarily draw on explicit knowledge.

While retrospective questionnaires are coarse instruments to assess awareness levels, they allow us to explore potential relationships between awareness levels and EIT scores. Erlam and Wei (2021) recently proposed a useful criterion for assessing the use of implicit knowledge during EITs. They hypothesized that, if EIT performance is based on implicit knowledge, it would not be impacted by the participants' awareness level. To test this assertion, Erlam and Wei administered an EIT with a plausibility judgement component to L2 English speakers in New Zealand, who were subsequently divided into "aware" and "unaware" groups based on their retrospective questionnaire responses. In line with their prediction, the authors found no significant EIT score difference between the two groups, suggesting that their EIT taps into implicit knowledge. In the current study, Erlam and Wei's hypothesis is further tested using a retrospective questionnaire to elucidate the role of awareness in EIT performance.

Comparison of performance on an EIT and other L2 tasks

In previous validation studies of tasks designed to measure explicit and implicit knowledge, *timed* tasks (including EITs) were proposed to assess implicit knowledge, whereas *untimed* form-focused tasks such as grammaticality judgement task (GJT)—were considered to tap into explicit knowledge. In an often-cited study conducted by R. Ellis (2009), three timed tasks (EIT, oral narrative task, and timed GJT) and two untimed tasks (untimed GJT and metalinguistic knowledge task) were administered to 91 L2 English speakers in New Zealand to tap into implicit and explicit knowledge, respectively. The confirmatory factor analysis results supported the author's hypothesis that the scores on the two task types loaded onto two separate factors (labeled as "implicit" and "explicit"). Based on this finding, R. Ellis argued that EITs have the potential to precisely gauge implicit knowledge of grammatical features of researchers' interest. These results were subsequently replicated by other authors who administered EITs to foreign language learners in non-immersion settings (e.g., Bowles, 2011; Zhang, 2015), lending further support for EITs as a measure of implicit knowledge.

Although these validation studies serve as an important initial step toward developing reliable measures of implicit knowledge, further work is still needed to resolve certain ambiguities. For instance, even if EIT scores correlate positively with other timed form-focused test scores (e.g., timed GJTs), this link may not necessarily be considered as strong evidence for EIT as a measure of implicit knowledge. Because timed GJTs intentionally direct test-takers attention to linguistic forms and grammaticality, it is possible that explicit knowledge can be deployed rapidly even under time pressure. It is thus critical to use fine-grained real-time grammar comprehension tasks that can direct participants' attention away from linguistic errors and indirectly assesses their implicit knowledge (Jiang, 2011; Suzuki, 2017; Vafaee, Suzuki, & Kachinske, 2017).

Word-monitoring tasks can be used for this purpose, as they measure sensitivity to grammatical errors while test-takers listen to a sentence in which the target errors are embedded immediately before the monitoring word (Jiang, 2011; Suzuki, 2017). For instance, test-takers could be presented with sentences with third person *s*, such as "The man in the library enjoy(s) reading difficult books." When participants listen for a monitoring word (e.g., reading) in an ungrammatical sentence, if they can detect the error, they would respond to the monitoring word more slowly than they would when presented with a grammatical sentence. Therefore, the reaction time (RT) difference between grammatical and ungrammatical items (defined as "grammaticality sensitivity index" or GSI) indicates the extent to which a decline in processing speed is caused by grammatical error detection. While this method has attracted some criticism (e.g., Godfroid & Kim, 2021), findings yielded by behavioral (Suzuki, 2017; Suzuki & DeKeyser, 2015; Vafaee et al., 2017) and neuroimaging experiments (Suzuki et al., 2022) support the construct validity of word-monitoring tasks as a measure of implicit knowledge. Therefore, in this study, performance on a word-monitoring task is used as the behavioral criterion measure of implicit knowledge.

Relative importance of explicit and implicit knowledge in the EIT performance

To our knowledge, there are only two previous studies in which the researchers compared participants' EIT scores with their implicit knowledge scores (derived from a word-monitoring task) as well as explicit knowledge scores (derived from a metalinguistic knowledge task). Suzuki and DeKeyser (2015) recruited 63 advanced Japanese L2 speakers with L1 Chinese who lived in Japan, all of whom performed an EIT to test their grammatical knowledge of five Japanese particles. Although EIT score was only weakly correlated with word-monitoring score ($r = 0.37$), a stronger correlation was found between the EIT score and the explicit (metalinguistic) knowledge score ($r = 0.46$). Based on the relative strengths of relationships between EIT performance and implicit and explicit knowledge task scores, Suzuki and DeKeyser argued that L2 speakers draw on explicit knowledge quickly during the EIT. Hence, the authors posited that EIT may be considered as a measure of automatized explicit knowledge. Automatized explicit knowledge refers to a body of knowledge that learners have conscious access to and can deploy quickly (Suzuki, 2017). It is also called "speeded-up" explicit knowledge, as it is not fully automatized or does not completely lack conscious awareness. As a result, automatized (speeded-up) explicit knowledge is distinct from implicit knowledge, but it is still highly functional and supports fluent communicative use of a L2.

In a more recent large-scale study, Godfroid and Kim (2021) administered nine grammatical knowledge tests targeting implicit and (automatized) explicit knowledge and four implicit aptitude tests to 131 L2 speakers at a university in the United States. Analyses of participants' scores on the EIT, word-monitoring, and metalinguistic knowledge tasks revealed a stronger correlation between the EIT and the explicit (metalinguistic) knowledge scores ($r = 0.34$) than the word-monitoring scores ($r = 0.08$). Although these patterns concurred with Suzuki and DeKeyser's (2015) results, based on their structural equation modeling (SEM) analysis including individual difference measures of aptitude, Godfroid and Kim (2021) concluded that EIT is a measure of implicit knowledge.[1] The divergent conclusions from these two studies may illustrate some limitations of behavioral evidence, as correlational patterns (including a more complex covariance matrix used in SEM) can only allow us to infer the knowledge (unobservable construct) underlying the measured variables but cannot directly reveal the neurocognitive processes recruited to demonstrate such knowledge. Thus, in order to complement these results and facilitate more accurate interpretations of behavioral findings, it is essential to employ neuroimaging techniques to investigate the brain regions associated with explicit and implicit knowledge.

Using neuroimaging techniques to study explicit and implicit knowledge

According to neurobiological models of L2 acquisition (see Fig. 1; Paradis, 2009; Ullman, 2020), explicit knowledge is encoded in declarative memory (i.e., a domain-general system for learning facts and events), whereas implicit knowledge is typically associated with procedural memory (i.e., a domain-general system for learning of motor and cognitive skills and forming habits). At the neural representation level, declarative memory is linked to hippocampus and medial temporal lobe structures. In particular, hippocampus is implicated in grammar learning when metalinguistic rules are provided to develop explicit knowledge (Tagarelli, Shattuck, Turkeltaub, & Ullman, 2019). On the other hand, procedural memory is primarily associated with frontal cortical-basal ganglia regions, which also play a key role in grammar learning. According to Ullman's (2020) prediction, the basal ganglia (particularly, the anterior caudate nucleus) are primarily recruited in the early phases of grammar learning, whereas frontal cortical regions, particularly in the premotor cortex (BA6) and the inferior frontal gyrus (IFG, BA44), become more important for the later stages of skill development (e.g., automatization).

[1] Further discussion on the SEM results is beyond the scope of this study.

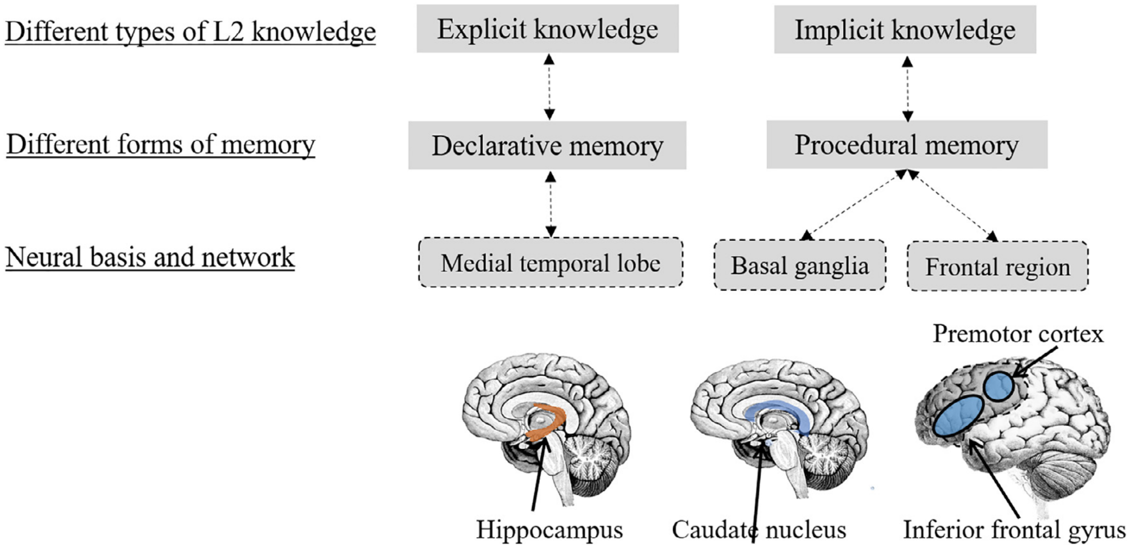

Fig. 1. Explicit–implicit knowledge, declarative–procedural memory, and neural network.

Table 1
Working Hypothesis on the EIT as a Measure of Explicit and Implicit Knowledge.

	EIT Components	
	Listening	Speaking
Knowledge Memory Systems and Brain Circuits	Implicit knowledge Procedural memory - Earlier developmental stage: Caudate nucleus - Later developmental stage: Left IFG and premotor cortex	(Speeded-up) explicit knowledge Declarative memory (e.g., hippocampus)

Based on these neurobiological models, an MRI can provide a promising means to reveal the knowledge types (explicit and implicit knowledge) underlying task performance by linking them to neural circuits (declarative and procedural systems). In the current study, we present and test our working hypothesis regarding the construct validity of EIT (see Table 1).

This hypothesis is based on the results yielded by the behavioral experiment conducted by Suzuki and DeKeyser (2015) indicating that the listening and speaking EIT phases may draw on explicit and implicit knowledge to different degrees. That is, implicit knowledge is primarily recruited during listening, whereas explicit knowledge is primarily used for the speaking phase. Because participants focus on meaning during the listening phase (as each stimulus sentence is followed by a comprehension question) and cannot know whether/when errors will occur, there is a limited opportunity to deploy explicit knowledge intentionally, given that they have only a few hundred milliseconds to process specific grammatical structures (e.g., Paradis, 2009; Suzuki, 2017). In line with this prediction, the neural processing during listening tasks may be based on procedural memory. This prediction was corroborated by the results obtained by Suzuki et al. (2022) in a recent fMRI study involving advanced L2 speakers, indicating that the procedural memory system (i.e., the left caudate nucleus and premotor cortex) was implicated during real-time listening comprehension of grammatical structures (assessed by a word-monitoring task).

In contrast, in the speaking phase, where several seconds are allowed for repeating the sentence, L2 speakers may be more likely to direct their attention to linguistic forms (in combination with meaning) and consciously monitor their utterances using speeded-up (automatized) explicit knowledge (e.g., Suzuki & DeKeyser, 2015). Hence, the role of declarative memory—implicated in the retrieval of explicit knowledge—should be prominent in the speaking component of EIT. This is our working hypothesis, which can be verified, modified, or rejected by fMRI findings, as this neuroimaging technique allows us to examine how neural circuits underlying explicit and implicit knowledge (i.e., declarative and procedural memory) are activated during the listening and speaking phases of EIT.

The current study

The goal of this study is to elucidate the extent to which EIT draws on explicit and implicit knowledge from both behavioral and neural perspectives. Advanced L2 Japanese speakers, as well as Japanese L1 speakers as controls, performed an EIT targeting Japanese case-markers inside the MRI scanner. A retrospective questionnaire was given immediately after the EIT to assess their awareness of ungrammatical items featured in the EIT. Metalinguistic knowledge (explicit knowledge) and word-monitoring (implicit knowledge) tasks were also administered to examine systematic relationships with EIT performance. Concerning the behavioral data, two research questions (RQs) were addressed:

1. To what extent is the awareness of grammatical errors related to EIT performance?
2. To what extent are explicit and implicit knowledge, measured by metalinguistic knowledge and word-monitoring tasks, associated with the EIT score?

RQ1 explores the relationship between the reported awareness of errors and EIT performance. If participants rely on implicit knowledge during the EIT, the scores achieved by “aware” and “unaware” groups (especially those on the ungrammatical items) will not be significantly different (Erlam & Wei, 2021).

RQ2 concerns the relationship between scores on the EIT and representative tests of explicit and implicit knowledge. Based on the findings reported by Suzuki and DeKeyser (2015) and Godfroid and Kim (2021), EIT scores might be positively related to those obtained on metalinguistic knowledge tasks, but the extent to which such a relationship can be established with word-monitoring scores remains uncertain.

Motivated by neurobiological models of declarative and procedural memory developed by Ullman and Paradis (Paradis, 2009; Ullman, 2020), three further RQs were addressed in this study to examine the neural processes involved in EIT performance:

3. What are the neural responses to grammatical errors during an EIT?
4. What is the relationship between awareness of grammatical errors and neural responses during an EIT?
5. To what extent are explicit and implicit knowledge, measured by metalinguistic knowledge and word-monitoring tasks, associated with the activations in declarative and procedural memory during an EIT?

RQ3 focused on the neural responses associated with the listening and speaking EIT phases. Based on our working hypothesis (Table 1), during the listening phase, test-takers may recruit implicit knowledge to detect ungrammaticality, whereas the speaking phase may allow for the use of (speeded-up) explicit knowledge. It was thus hypothesized that, in the listening process, the procedural memory network (premotor cortex, left IFG, and caudate) is preferentially recruited, whereas the speaking process preferentially engages the declarative memory system (e.g., hippocampus).

Regarding RQ4, we hypothesized that L2 speakers with higher awareness would rely on their procedural memory during the listening EIT phase more strongly than those with no/lower awareness. This differentiation was based on the view that L2 speakers with sufficient implicit knowledge would be more likely to detect errors in real-time comprehension. As a result, they would be more aware of errors in the EIT upon its completion (Suzuki et al., 2022). Put differently, L2 speakers with little implicit knowledge cannot detect errors while listening, as their attention is primarily directed to meaning, which leads to lower levels of awareness. There are two possible predictions for the speaking phase. On the one hand, L2 speakers that exhibit greater awareness may continuously rely on implicit knowledge during the speaking EIT phase. On the other hand, it is also possible that they may rely on explicit knowledge as they are more likely to be alerted to grammatical errors.

Regarding RQ5, we hypothesized that implicit knowledge (word-monitoring) score would correlate with the activation of procedural memory (the frontal-basal ganglia circuits) during the listening phase, whereas metalinguistic knowledge task performance would be related to the neural circuits associated with declarative memory (hippocampus) during the speaking phase.

Method

Participants

Participants for this study were recruited at a national university located in the northern part of Japan. Only the individuals that met the following inclusion criteria were eligible for participation: (a) native Mandarin speakers, (b) advanced Japanese proficiency equivalent to N1 in the standardized Japanese Language Proficiency Test (JLPT), which is the minimum requirement for acceptance into a regular college undergraduate/graduate program in Japan, (c) arrived to Japan at the age of 17 or older, and (d) living in Japan for at least 12 months.

Although 32 L2 Japanese speakers met these stringent requirements and were enrolled in this study, data pertaining to seven participants were subsequently excluded from the analyses, due to one of the following reasons: (a) absence in one of the experimental sessions ($n = 4$), (b) an experimenter error ($n = 1$), and (c) excessive motion exceeding 3 mm within the scanner ($n = 2$). Data related to the remaining 25 participants (10 males, 15 females) was analyzed and is reported in this paper. The participants' background information is presented in Table 2. In terms of their academic level, they were undergraduate ($n = 3$), research ($n = 3$), master's ($n = 17$), and doctoral ($n = 1$) students. More than half of participants ($n = 14$) obtained Bachelor's degree in Japanese as a major at a Chinese university, while other participants obtained Bachelor's degree in other fields (e.g., biology, engineering, food science, environment). In addition, four participants were pursuing or had obtained Master's degree in Japanese linguistics at a Japanese university.

To establish the baseline for L1 neural responses during the EIT, 21 native Japanese speakers were also recruited from undergraduate courses offered by the same university (14 males, 7 females; mean age = 21.57 years, $SD = 1.62$, range: 18–24).

All participants met the fMRI experiment requirements, as they were right-handed, of normal hearing, had either normal or corrected-to-normal vision, and did not suffer from any neurological deficits or psychiatric disorders. This study was conducted with the approval of the Institutional Review Board of the university from which the study participants were recruited. Written informed consent was obtained from each participant prior to the experiment.

Table 2
Background Information for L2 Speakers.

	M	SD	Min	Max
Age	24.25	1.70	20	27
Starting age of instruction	18.21	1.47	15	21
Length of instruction (months)	50.29	27.28	6	120
Age of arrival	21.42	1.61	18	24
Length of residence (months)	30.17	12.64	12	58

Note. One participant failed to complete the questionnaire.

Target structures

Four grammatical structures were used for this study: (a) case-marking particles *o-ga* for transitive-intransitive verb pairs, (b) case-marking particles *wa-ga* in adverbial clause, (c) case-marking particles *wa-ga* in relative clause, and (d) locative particles *ni-de*. These structures were used by Suzuki and DeKeyser (2015) in their EIT validation study and are usually taught explicitly in Japanese classes. In the debriefing questionnaire, all learners reported to have studied about the transitive-intransitive verbs and *ni-de* in school and/or grammar reference books, while some of them indicated no recollection of having learned about *wa-ga* in adverbial clause (eight learners) and *wa-ga* in relative clause (four learners).

Particles o-ga for transitive-intransitive verbs. Sixteen transitive/intransitive verb pairs that share the stem and morphological markings that differentiate transitive from intransitive verbs were chosen for the task. Example (1a) illustrates a sample grammatical and ungrammatical sentence with a transitive verb (*akeru*, "open"). A theme (*mado*, "window") should be followed by the object-marking particle *o* rather than the subject-marking particle *ga*. In contrast, as shown in Example (1b) with an intransitive verb (*hajimaru*, "start"), the subject should be followed by the subject-marking particle *ga* rather than *o*.

(1a) *Fuyu ni mado o/*ga <u>akeru</u> to, samui to omou.*
Winter window–OBJECT open if, cold that think.
I think it is cold if we open the window in winter.
(1b) *Natsuyasumi ga/*o <u>hajimaru</u> to, gakusei wa ureshii.*
Summer vacation–SUBJECT start if, students–SUBJECT happy.
When summer vacation starts, students become happy.

Particles wa-ga in adverbial clause. L2 Japanese learners are often confused by topic-marking (*wa*) and subject-marking (*ga*) particles. One of the distinctions made between the case-marking particles *wa* and *ga* is based on their location in the sentence structure. *Ga* should be used rather than *wa* within the adverbial clause, as illustrated in Example (2).

(2) *Chugakusei ga/*wa tabako o sutteita ra, <u>otona</u> wa okoru bekida.*
JHS students–SUBJECT smoke if, adults–SUBJECT scold should.
If a junior high school student is smoking, adults should scold.

Particles wa-ga in relative clause. In a similar vein, the case-marking particle *ga* should be used rather than *wa* within relative clauses, as illustrated in Example (3).

(3) *Yumeijin ga/*wa sumu <u>manshon</u> wa takai darou.*
Celebrity–SUBJECT live mansion–TOPIC expensive maybe.
The mansion in which celebrity lives may be expensive.

Particles ni-de. The locative particles *ni* and *de* are distinguished by the verb semantics. *De* should be used for indicating the location at which an action takes place, while *ni* is mainly used for stative verbs (e.g., be, live). Example (4) illustrates this restriction with an action verb (*kaimonosuru* "do shopping").

(4) *Konbini de/*ni <u>kaimonosuru</u> no wa totemo benri da.*
Convenience store–LOC do shopping TOPIC very convenient be.
It is convenient to do shopping at the convenience store.

Instruments

EIT. Fig. 2 illustrates the EIT procedure in a 3T-fMRI scanner. Adopting an event-related design, a fixation cross first appeared for 1 second, followed by an auditory sentence. After the sentence ended, a yes-no semantic plausibility judgement question appeared on the screen. When participants made a plausibility judgement, a number immediately appeared on the screen, and the participants were required to count aloud the numbers in Japanese (*san*, *ni*, *ichi* [three, two, one]), which prevented rote repetition. After the prompt, they were given 8 s to repeat the sentence, in accordance with the procedure adopted by Suzuki and DeKeyser (2015). Once the 8 second period lapsed, the trial terminated, and the next trial started after a short resting period of 2 – 10 second duration which was inserted between consecutive trials to increase efficiency, allowing random sampling of hemodynamic response function (HRF) for imaging data.

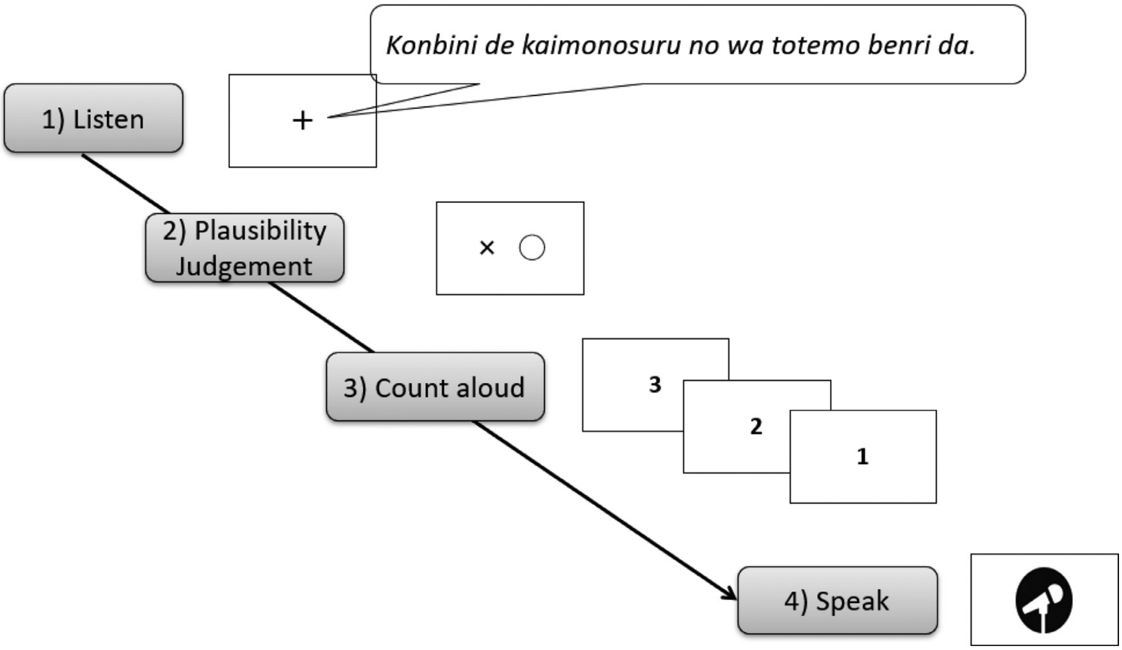

Fig. 2. EIT procedure.

Prior to commencing the test, participants were instructed (a) to judge whether or not the sentence they heard was semantically plausible and (b) to repeat it in correct Japanese within eight seconds. For the latter, participants read a note saying "(If necessary) repeat in a grammatically correct sentence" with a written example as follows:

A sentence you hear: *Steve Jobs wa yumeikatta*. (Steve Jobs was famous.)
Repetition: *Steve Jobs wa yumeidatta*. (Steve Jobs was famous.)

This example illustrated that the sentence participants would hear might contain an ungrammatical aspect (i.e., wrong conjugation of adjective) that was not relevant to the target grammatical structure in the main EIT. Because some participants might try to intentionally repeat ungrammatical sentences as they were (Spada et al., 2015; Suzuki & DeKeyser, 2015), explicitly instructing participants to correct ungrammatical sentences was necessary (see Limitations and Suggestions for Future Research). To gain familiarity with the EIT procedure, participants also engaged in a practice session comprising 10 items that did not include the target grammatical structures (half plausible and half implausible) outside and inside the fMRI scanner.

The experimental protocol consisted of 96 trials, 64 of which were critical trials (plausible sentences) and 32 were filler trials (implausible sentences). The critical trials included 32 grammatical (8 sentences × 4 structures) and 32 ungrammatical sentences. Two counterbalanced lists were created for the 64 critical trials (see Appendix A in Online Supplementary File for details). The 32 grammatical sentences in List 1 had corresponding ungrammatical sentences in List 2, and vice versa. They were divided into "repeat" trials and "no-repeat" trials in which participants were not required to repeat the sentence (i.e., the system moved on to the next trial after the prompt). The participants' speech on the "repeat" trials was analyzed for speech production. To differentiate brain responses during listening from those during speaking, the data gathered during the listening phase in both "no-repeat" and "repeat" trials was analyzed.

First, accuracy of the plausibility judgement questions was computed.[2] The mean accuracy rate was high for both L1 and L2 speakers (L1: $M = 96.44\%$, $SD = 6.13\%$; L2: $M = 93.82\%$, $SD = 5.25\%$). In the study conducted by Suzuki and DeKeyser (2015), data pertaining to the participants whose accuracy was below 85% was excluded from the analysis, and only one L2 speaker in the current study scored below that cutoff point (77%). The data related to this participant was included in the current analyses because performing the EIT inside an fMRI scanner was deemed more difficult. Next, two trained coders scored participants' production performance in terms of three criteria (e.g., Erlam, 2006): (a) obligatory occasion created – required form supplied; (b) obligatory occasion created – required form not supplied; and (c) no obligatory occasion created.[3] A credit was only given for the first category, while the remaining two categories were scored as incorrect. Utterances with self-correction within the response time were also scored as correct. Reliability indexed by Cronbach's alpha was acceptable for the two counterbalanced lists (List 1 = 0.83 and List 2 = 0.87).

[2] Implausible sentences were to be judged accurately based on common sense (e.g., *Basukettobooru o suru toki wa, ashi de booru o takusan keru*, "When playing basketball, we kick the ball a lot").

[3] Because L1 speakers correctly responded to the comprehension question and indicated compliance with the current EIT procedure, L1 speech data (presumably reaching the ceiling performance) were not further coded.

Retrospective questionnaire. A recall questionnaire was administered to all participants immediately after EIT completion. In the main part of the questionnaire, the participants were asked whether they noticed any errors while listening (e.g., Granena, 2016). If they responded affirmatively, they were prompted to provide examples of sentences containing grammatical errors. Based on their responses, the participants were categorized to three awareness groups: (a) no awareness, (b) lower-level awareness (those who reported noticing errors but could not provide any of the targeted rules), and (c) higher-level awareness (those who reported noticing errors and provided at least one of the targeted rules).

Word-monitoring task. The word-monitoring task was completed in the MRI scanner. Participants (a) saw a monitoring word (which remained on the screen until it was identified by the button press), (b) listened to a sentence for that monitoring word and pressed the button as soon as they identified it in the sentence, and (c) made a semantic plausibility judgement of the sentence. The list of stimulus sentences was counter-balanced with that of the EIT. The monitoring word was always a content word and they are underlined in Example (1)–(4) above. GSI was computed by subtracting grammatical RT from ungrammatical RT, indicating online sensitivity to grammatical errors (e.g., Godfroid & Kim, 2021; Suzuki & DeKeyser, 2015; Suzuki, 2017). Detailed behavioral and brain analysis results are reported by Suzuki et al. (2022). Reliability indexed by Cronbach's alpha for the word-monitoring task was high for the two counterbalanced lists (List 1 = 0.96 and List 2 = 0.98).

Metalinguistic knowledge task. After the word-monitoring task, participants took a paper-and-pencil metalinguistic knowledge task, which consisted of (a) a correction and (b) an explanation component. They were told that each sentence contained one grammatical error and were instructed (a) to underline the part where they believe the grammatical error exists and write down the correct Japanese term below, and (b) to explain why the original was incorrect (either in Japanese or Chinese). The stimulus list contained 16 ungrammatical sentences (4 sentences × 4 target structures), all of which were extracted from the stimulus list for the EIT. No time limit was imposed for the completion of this task. Two native Japanese speakers used an appropriate rubric to independently score the responses, achieving 98.25% inter-rater reliability, with the remaining inconsistencies resolved by a third coder. A credit was given only when the participant provided both the correction and the explanation of the target rule. Reliability indexed by Cronbach's alpha was 0.86.

Procedure. Participants attended two test sessions in the laboratory, completing either the EIT or the word-monitoring task in the MRI scanner during the first session, with the other task completed in the second session using a counter-balanced design. At the end of the second session, the metalinguistic knowledge task was administered outside the scanner in a quiet room. This order minimized the potential influence of taking the metalinguistic knowledge task on the more implicit tasks. All materials will be made available in IRIS Digital Repository (Marsden, Mackey, & Plonsky, 2016).

Statistical analyses

Behavioral data analysis (RQ1 and RQ2). First, to examine the relationship between awareness and EIT performance, independent-samples t-tests were conducted on the EIT data to compare the performance of (a) participants who showed higher level of awareness with that of (b) others who showed lower level of or no awareness (RQ1). According to the L2-specific research benchmark (Plonsky & Oswald, 2014), the magnitude of effect size (d) was interpreted as small (0.40), medium (0.70), or large (1.00). Second, multiple regression analysis was conducted on the EIT score as a dependent variable with the metalinguistic knowledge and word-monitoring scores as predictors (RQ2). In order to examine the relative importance of predictors, dominance analysis was further conducted using a free R-based web application (https://langtest.jp/shiny/relimp/). Because standardized beta coefficients from multiple regression can sometimes be misleading due to the suppression effect of predictors, dominance analysis facilitates the evaluation of relative importance of predictors (see Mizumoto, 2022 for details). Specifically, dominance weights are calculated to compare the relative effect sizes of predictors. According to the Shapiro-Wilk test, all dependent and independent variables were normally distributed ($p > .05$).

Brain data analysis (RQ3–5). The functional imaging data of 25 L2 learners were analyzed for event-related fMRI signals (see Appendix B in Online Supplementary File for details). 21 L1 speakers' data were also analyzed because their linguistic knowledge and skills are highly automatized and presumably implicit, and their neural responses will thus serve to locate the brain regions associated with the procedural system engaged during the EIT. In the first-level analysis, the degree of activation was estimated based on a voxel-by-voxel multiple regression analysis of the time courses. A general linear model was constructed for each participant to analyze the hemodynamic responses captured by functional images. Six regressors were modelled for three stimulus types (grammatical, ungrammatical, and fillers) during the listening and speaking phases. In addition, if participants provided an incorrect answer to the plausibility judgement question and/or incorrect sentence during the speaking phase, those trials were separately modeled as error. Six movement parameters (three translations, three rotations) were also included as regressors of no interest. A high-pass filter with a 128 second cut-off period was used to eliminate any artifactual low-frequency trends. Each trial was modeled as an epoch for the duration of auditory sentence in the listening phase and duration of repetition in the speaking phase. Then, the contrast images for ungrammatical > grammatical sentences in each listening and speaking component were generated for the L2 and L1 groups, respectively. The statistical threshold was set at $p < .05$ with a cluster-level family-wise error (FWE) correction (initial height voxel threshold, $p < .001$) for whole-brain analysis. Moreover, for regions with a priori hypotheses, small volume correction (SVC) with anatomical mask of each area was applied for voxel-level multiple comparisons ($p < .05$ with FWE correction).

Three RQs were addressed by conducting the following inferential statistical tests on the brain data obtained during the listening and speaking EIT phases. First, to answer RQ3, statistical inference was performed on parameter estimate contrasts in a random effects second-level analysis, whereby one-sample t-tests were conducted for each group and two-sample t-tests were utilized for comparison between L2 and L1 groups.

Table 3
Descriptive Statistics for L2 Speakers.

	M	SD	Min	Max
EIT				
Grammatical Items (%)	78.90	16.40	37.50	100
Ungrammatical Items (%)	67.50	20.70	25.00	100
Total (%)	73.20	17.20	37.50	100
Metalinguistic Knowledge (%)	65.00	16.94	37.50	100
GSI – Word-Monitoring (ms)	7.48	51.48	−130	103

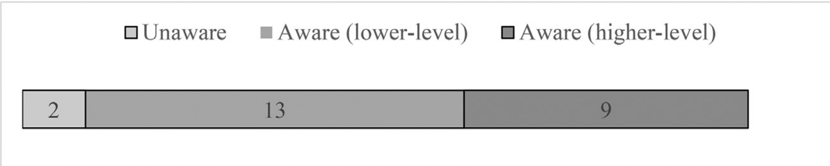

Fig. 3. Reported awareness of grammatical errors in the EIT.
Note. One participant failed to complete the questionnaire.

Second, to examine the relationship between awareness level and neural activity during the EIT (RQ4), two-sample t-tests were conducted on the contrasts of ungrammatical > grammatical sentences to compare L2 subgroups (higher-awareness group versus no + lower-awareness group) at the whole brain level separately for the listening and speaking phases.

Third, multiple regression and dominance analyses were conducted on the brain activity during the EIT with metalinguistic knowledge and word-monitoring scores as predictors (RQ5). In these models, brain activity in regions of interest (ROIs) served as dependent variables based on the neurobiological framework of declarative and procedural memory (see Fig. 1). Specifically, the neural responses to the contrast of ungrammatical > grammatical items were computed separately for the listening and speaking phases in the following four ROIs: three functional ROIs for procedural memory (premotor cortex, left IFG, left caudate nucleus) and one anatomical ROI for declarative memory (left hippocampus). The ROI for the premotor cortex (x, y, z = −40, −3, 32) was taken from Suzuki et al.'s (2022) study, as their findings showed that L1 speakers exhibited heightened activity in this region to detect grammatical errors during the word-monitoring task. The ROI for the left IFG (x, y, z = −52, 8, 4) was found to be significantly activated in the current L1 group who performed the same EIT (see the Results section). The ROI for the left caudate was chosen based on Suzuki et al.'s (2022) study, which demonstrated the heightened activation in the left head of caudate nucleus (x, y, z = −4, 10, 6) during the word-monitoring task. Finally, the CA1–3 regions of hippocampus was selected as the most appropriate anatomical ROI using the Anatomical toolbox (Eickhoff et al., 2005). This region was chosen based on a recent 7T MRI neuroimaging finding that CA1–3 regions of the hippocampus are particularly related to explicit memory during both information encoding and retrieval (e.g., Seok & Cheong, 2020). To clarify the activation profile in a particular brain area, we extracted parameter estimates in each condition for each participant using the Marsbar toolbox (Brett, Anton, Valabregue, & Poline, 2002).

Results

Behavioral results

Descriptive statistics. Table 3 presents descriptive statistics.

Role of awareness in EIT (RQ1). All except for two participants (22 out of 24) reported that they noticed grammatical errors in the stimulus sentences (see Fig. 3).[4] However, only nine of these individuals were able to correctly recall at least one targeted grammatical error embedded in the sentence.[5]

In order to compare the participants who showed higher-level awareness (n = 9) with the remaining cohort, those with no and lower-level awareness were placed into one group (n = 15) before conducting an independent-samples t-test.[6] As shown in Fig. 4, the higher-level awareness group obtained significantly higher scores than the no + lower-level awareness group with large effect sizes: total score ($p < .01$, d = 1.36, 95% CI [0.43, 2.27]), grammatical items ($p < .01$, d = 1.19, 95% CI [0.28, 2.15]), and ungrammatical items (p = .01, d = 1.19, 95% CI [0.28, 2.07]).

[4] All 21 L1 speakers reported noticing grammatical errors, and 15 of these individuals correctly recalled at least one targeted grammatical error.

[5] The types of linguistic structures that were correctly reported were *ni/de* (6/23 participants), transitive-intransitive (4/23), adverbial clause (2/23), and relative clause (2/23). Other linguistic issues that the participants reported included the usage of adjective, existence verbs (*iru/aru*), verb conjugations, and other case-marking particles (e.g., objective marking particle such as *o*). In some cases, these participants provided some ungrammatical examples that they had presumably misheard in the stimuli.

[6] The two "unaware" participants showed below-average EIT performance (63% and 72%), and their accuracy on the ungrammatical items (50% and 63%) was lower than on the grammatical items (75% and 81%).

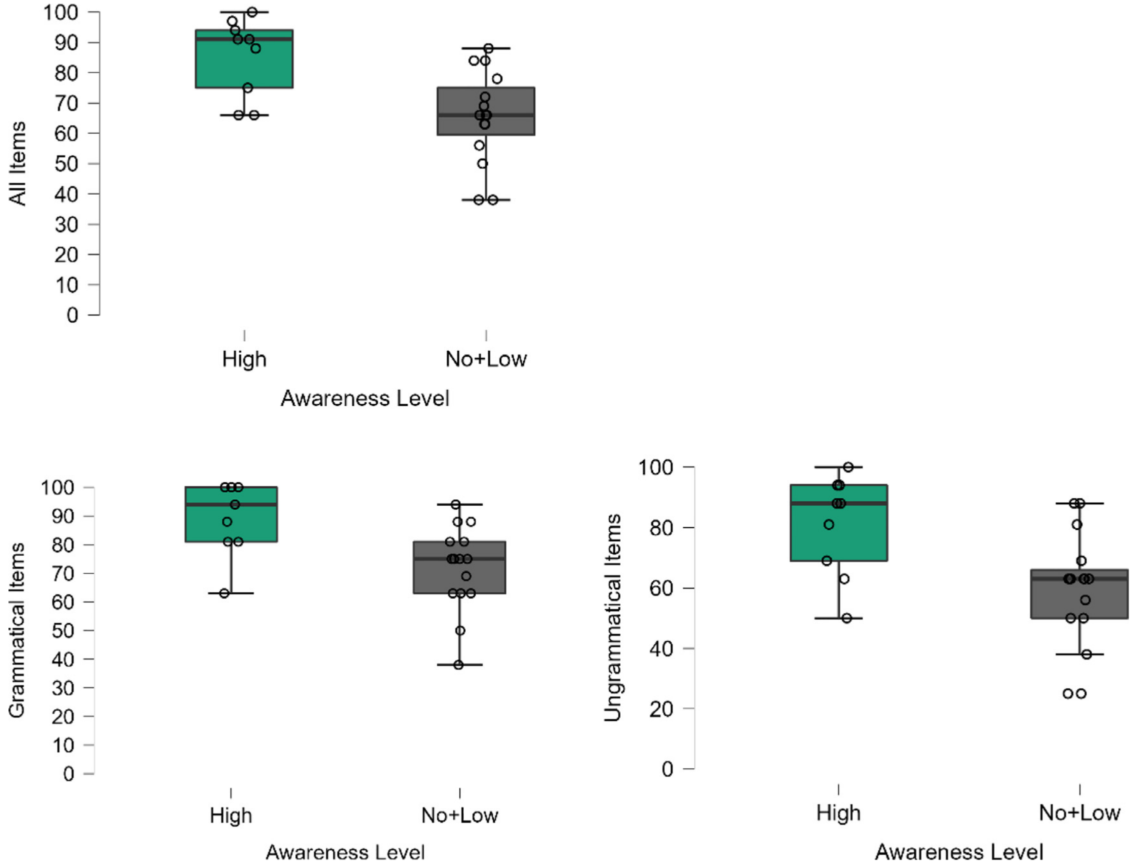

Fig. 4. EIT performance comparison between the higher-level awareness and the no + lower-level awareness groups. *Note.* The numerical values are presented in Appendix C in Online Supplementary File.

Table 4
Relative Importance of Word-Monitoring and Metalinguistic Knowledge Scores on the EIT Performance.

	Relative Importance (%)	95% CI	Estimate	SE	*t*	*p*
(Intercept)			0.36	0.10	3.47	0.00
Metalinguistic Knowledge	0.35 (69%)	[0.05, 0.55]	0.02	0.00	3.59	0.00
Word-Monitoring (GSI)	0.16 (31%)	[0.04, 0.32]	0.00	0.00	2.10	0.05

Note. The % value in the bracket indicates the proportion of explained variance in the entire regression model (hence, the sum of predictors is 100%). The correlation between metalinguistic knowledge and GSI was weak ($r = 0.26$, $p = .21$). Multiple regression results for grammatical and ungrammatical test items are presented in Appendix D in Online Supplementary File.

Relative importance of word-monitoring and metalinguistic knowledge scores (RQ2). Multiple regression analysis was conducted to investigate the relative importance of metalinguistic knowledge and word-monitoring (GSI) scores for the EIT score. The omnibus model was significant, accounting for 51% of variance in the EIT score, $F(2, 22) = 11.41$, $p < .001$, $R^2 = 0.51$. Both predictors were significant ($p < .05$). As shown in Table 4, the dominance analysis indicated that metalinguistic knowledge score and GSI explained 35% and 16% of the variance, respectively.

Brain results

Whole-brain analysis (RQ3). During the listening phase, the left opercular and triangular parts of IFG (posited to relate to procedural memory), as well as left insula, were significantly activated in the L1 group (see Appendix E in Online Supplementary File for details). In contrast, no significant activation in the L2 group was observed for the contrast of ungrammatical > grammatical items. As shown in Fig. 5, significantly greater activation in the opercular part of inferior frontal gyrus (BA44) was found in the L1 group relative to the L2 group (MNI *x, y, z* coordinates = −52, 8, 4, $t = 5.64$, 511 voxels, $p < .05$ with FWE correction, cluster level). In the speaking

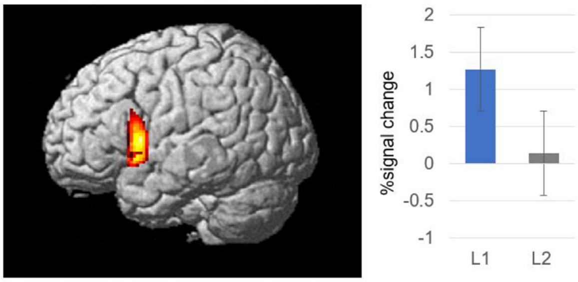

Fig. 5. Left IFG (L1 Speakers > L2 Speakers).

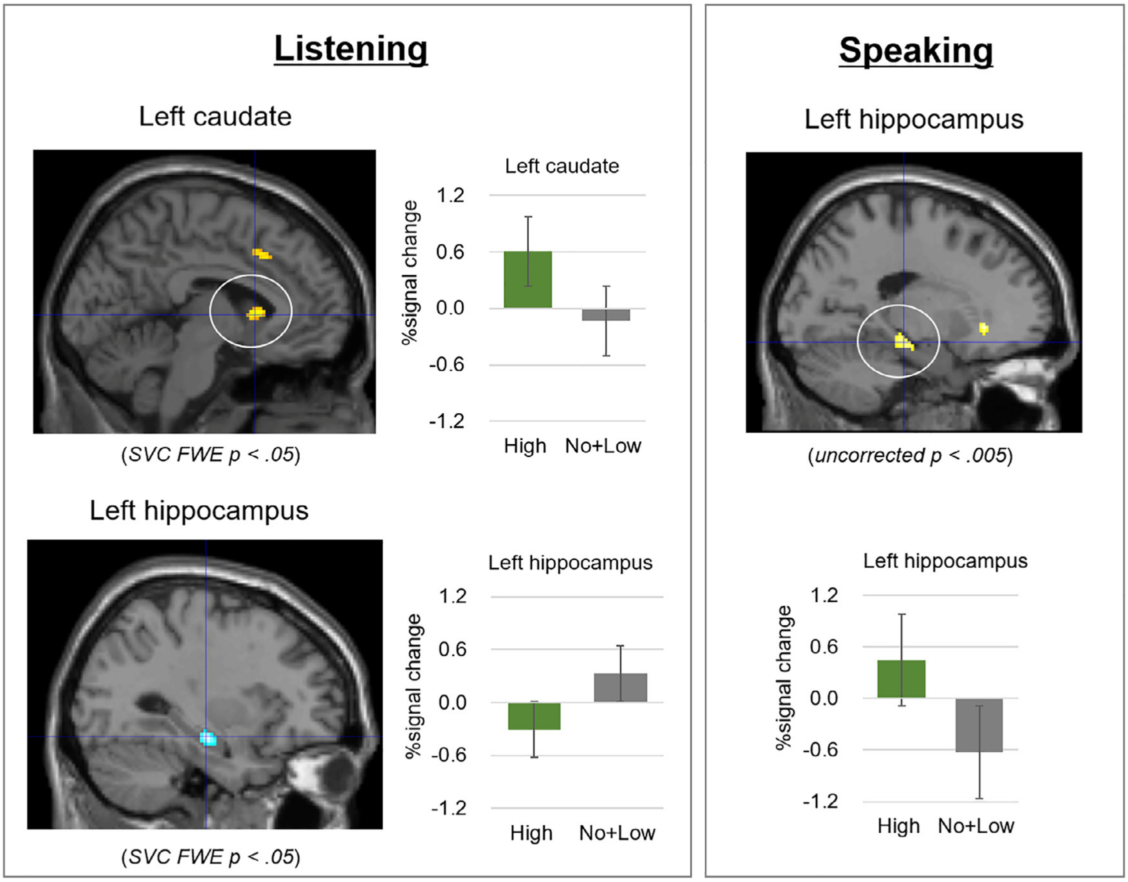

Fig. 6. Brain regions that showed differential activation between the higher-level awareness group and ($n = 9$) and no + lower-level awareness group ($n = 15$).

phase, there was no significant activation in the contrast of ungrammatical > grammatical items in either L1 or L2 group, suggesting that none of the brain regions were activated preferentially for either group.

Role of awareness (RQ4). Two-sample t-tests were conducted on the neural responses during the listening phase at the whole brain level. Fig. 6 illustrates a contrasting pattern between procedural and declarative systems recruited during the listening and speaking phases. During the listening phase, in L2 speakers with higher-level awareness, the left caudate nucleus exhibited significantly higher activation relative to those with no or lower-level awareness (MNI x, y, z coordinates = −4, 14, 4, $t = 4.06$, SVC, $p < .05$ with FWE correction). In contrast, significantly *lower* activation was noted in the left hippocampus (MNI x, y, z coordinates = −30, −16, −14, $t = 4.47$, SVC, $p < .05$ with FWE correction) in the higher-level-awareness group. During the speaking phase, the left hippocampus exhibited higher activation for L2 speakers with higher awareness relative to the L2 speakers with no + lower awareness, with the liberal statistical threshold (MNI x, y, z coordinates = −18, −24, −14, $t = 3.20$, $p < .005$ uncorrected).

Relative importance of word-monitoring and metalinguistic knowledge scores (RQ5). Multiple regression results related to the four ROIs are summarized in Table 5. The model for left IFG activation during listening was significant and accounted for 32.40% of the

Table 5
Multiple Regression Results.

	Premotor		LIFG		Left Caudate		Left Hippocampus	
	Listen	Speak	Listen	Speak	Listen	Speak	Listen	Speak
F	2.06	0.63	5.27	0.56	1.73	0.61	2.82	3.09
p	0.15	0.54	0.01	0.58	0.20	0.55	0.08	0.07
R^2	15.75%	5.40%	32.40%	4.86%	13.58%	5.26%	20.42%	21.93%

Table 6
Relative Importance of Word-Monitoring and Metalinguistic Knowledge Scores for Neural Activity.

[Left IFG – Listening]

	Relative Importance (%)	95% CI	Estimate	SE	t	p
(Intercept)			−1.63	0.56	−2.94	.01
Metalinguistic Knowledge	0.32 (98%)	[0.11, 0.55]	0.08	0.03	3.24	.00
Word-Monitoring (GSI)	0.01 (2%)	[0, 0.02]	0.00	0.00	−0.61	.55

[Left Hippocampus – Listening]

	Relative Importance (%)	95% CI	Estimate	SE	t	p
(Intercept)			0.67	0.27	2.46	.02
Metalinguistic Knowledge	0.19 (95%)	[0.01, 0.46]	−0.03	0.01	−2.37	.03
Word-Monitoring (GSI)	0.01 (5%)	[0, 0.05]	0.00	0.00	0.77	.45

[Left Hippocampus – Speaking]

	Relative Importance (%)	95% CI	Estimate	SE	t	p
(Intercept)			−1.23	0.42	−2.92	.01
Metalinguistic Knowledge	0.21 (94%)	[0.01, 0.46]	0.05	0.02	2.48	.02
Word-Monitoring (GSI)	0.01 (6%)	[0, 0.07]	0.00	0.00	−0.86	.40

Note. The% value in the bracket indicates the proportion of explained variance in the entire regression model (hence, the sum of predictors is 100%).

EIT score variance ($p < .01$). Two models for the left hippocampus activation were marginally significant for listening and speaking phases, both accounting for about 20% of the variance. Dominance analysis results are presented only for these three models (see Appendix F in Online Supplementary File for all dominance analysis results).

As shown in Table 6, metalinguistic score was a significant positive predictor of left IFG (listening) activation. In contrast, for the left hippocampus (listening), metalinguistic score was a significant *negative* predictor. During the speaking EIT phase, metalinguistic score was a significant positive predictor of left hippocampus activation. Overall, the importance of metalinguistic knowledge was dominant (94–98%), leaving only negligible effect of word-monitoring score (2-6%). In order to illustrate these significant patterns for these two brain regions, scatterplots relating metalinguistic scores and activity levels are provided in Fig. 7.

Discussion

Behavioral evidence

Although capturing the state of awareness in situ during an EIT is difficult using the post-task questionnaire, nine out of 24 L2 learners that took part in this study accurately reported at least one of the targeted ungrammatical structures embedded in the EIT. In answer to RQ1, these L2 speakers with higher-level awareness attained higher EIT scores compared to those with no + lower awareness. This pattern is inconsistent with the recent results obtained by Erlam and Wei (2021), suggesting that some of the L2 speakers that participated in our study, who noticed errors in certain target structures while engaging in the EIT, might have accessed their explicit knowledge in a more strategic, conscious manner and reformulated more accurate sentences.

Regarding RQ2, we hypothesized that, when taking the EIT, L2 learners draw on explicit knowledge more strongly than on implicit knowledge. Our multiple regression and dominance analysis findings revealed that, while their EIT performance was significantly predicted by both metalinguistic knowledge and word-monitoring scores, the contribution of the former was twice that of the latter. We thus replicated the findings reported by Suzuki and DeKeyser (2015) whose study targeted similar groups of participants (advanced L2 Japanese speakers with L1 Chinese living in Japan) and relied on similar target structures (case-markers). Guided by their results, Suzuki and DeKeyser (2015) argued that advanced L2 speakers can quickly retrieve explicit knowledge required for EIT performance. According to skill acquisition theory, this explicit knowledge is presumably acquired consciously and is strengthened/consolidated/automatized with practice (DeKeyser, 2017; Suzuki, 2022) so that it would be retrieved efficiently during EITs.

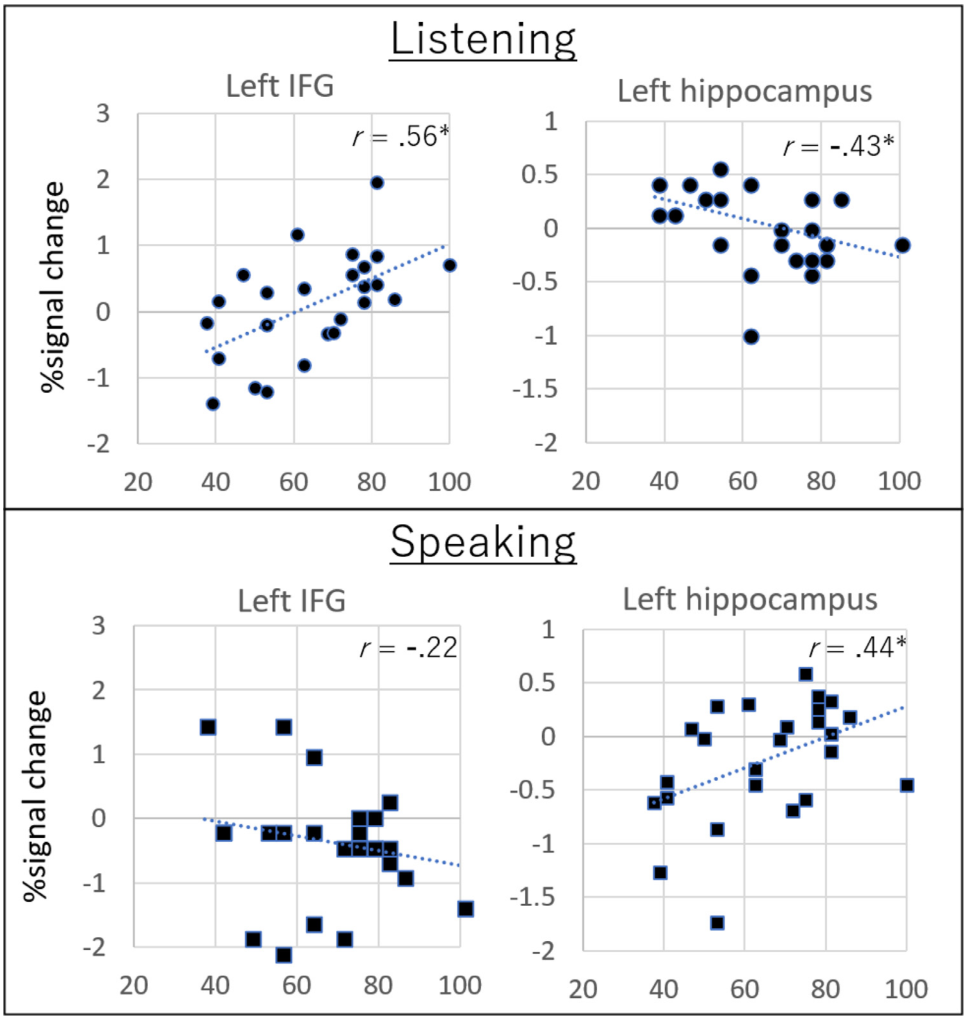

Fig. 7. Metalinguistic knowledge score scatterplots in relation to the left IFG and left hippocampus activation levels.

Based on this finding, it seems unrealistic to shut off access to speeded-up explicit knowledge, even if a semantic judgement component (directing attention to meaning) and time pressure are included in the EIT. Taken together, such behavioral evidence indicates that, for this group of advanced L2 users, speeded-up explicit knowledge is instrumental (and arguably more important than implicit knowledge) for comprehension and production skills, as measured by EIT.

Neural evidence

The first goal of neural analyses conducted in this work was to reveal the neural circuits activated when responding to ungrammatical and grammatical stimuli as a part of the EIT (RQ3). The whole-brain analysis revealed that L1 speakers recruit mainly procedural memory (i.e., left IFG) during the listening phase. For the L2 speakers, however, no significant activation of specific brain areas was detected for the grammatical vs. ungrammatical contrasts during either the listening or the speaking phase. This L2 whole-brain result did not support our working hypothesis (see Table 1) that L2 learners may recruit procedural/implicit knowledge (i.e., caudate nucleus, left IFG, and/or premotor cortex) for listening comprehension, whereas the speaking phase may allow for the rapid use of declarative/explicit knowledge (i.e., hippocampus). However, this gap between the hypothesis and the experimental findings was most likely due to greater variations among L2 speakers relative to L1 speakers, as discussed below.

When we compared the neural responses of L2 speakers with higher-level awareness (who accurately reported grammatical errors) with those of participants with no or lower-level awareness (RQ4), the findings were more aligned with the hypothesis. The higher-level awareness subgroup showed *stronger* left caudate activity but *weaker* hippocampus activity during the listening phase. In contrast, this same subgroup of learners exhibited *stronger* declarative system (i.e., hippocampus) activation during the speaking phase. As this pattern was confirmed only with the liberal statistical threshold in the analysis involving data pertaining to a small number of participants, extra caution is needed when interpreting this finding, which should be further attested through future in-

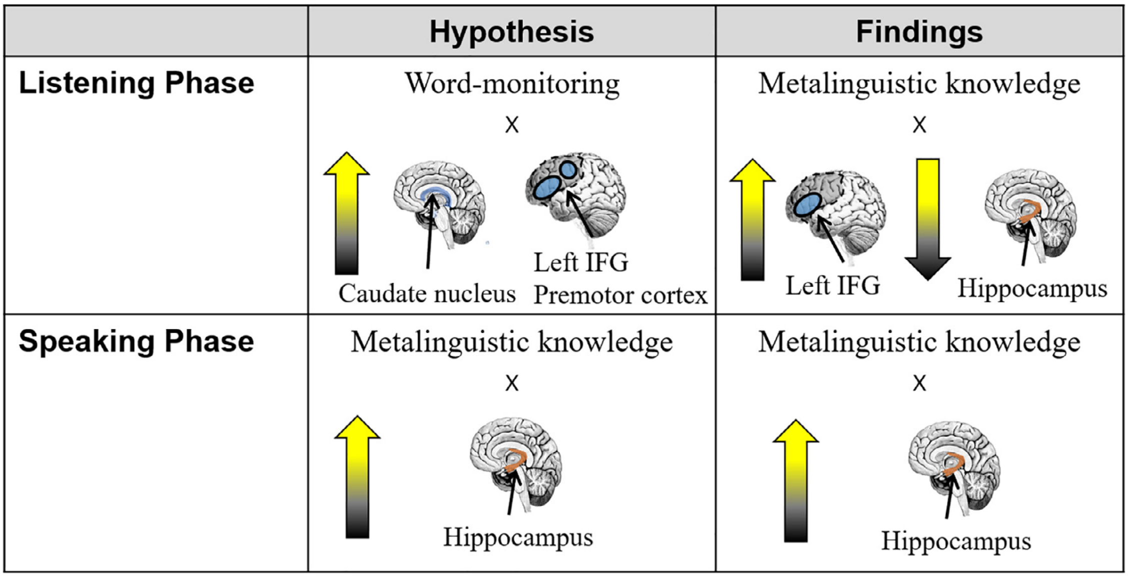

Fig. 8. Hypotheses and findings.

vestigations. However, this result suggests that some advanced L2 speakers with some immersion experience (a) detect grammatical errors spontaneously while listening primarily using implicit knowledge, (b) become aware of the errors after rehearsal in working memory, and (c) then consciously monitor their utterances primarily using explicit knowledge (see Suzuki & DeKeyser, 2015, p. 865 for this interpretation).

Finally, we explored the extent to which metalinguistic knowledge and word-monitoring task scores are associated with neural responses during the EIT (RQ5). The hypotheses and current findings are summarized in Fig. 8. Our hypotheses were straightforward, postulating that implicit knowledge (word-monitoring) scores would correlate with greater procedural memory activation during the listening phase, whereas (metalinguistic) explicit knowledge scores would be reflective of greater declarative memory activation during the speaking phase. The hypothesis was supported for the speaking phase during which left hippocampus activity (presumably associated with retrieval of information from declarative memory; Seok & Cheong, 2020) was positively correlated with the metalinguistic knowledge score. This neural evidence corroborates the behavioral finding that EIT taps into speeded-up explicit knowledge. Furthermore, activation of the brain regions associated with procedural memory (caudate nucleus, left IFG, and premotor cortex) was not systematically related to word-monitoring or metalinguistic knowledge scores, suggesting that implicit knowledge was not systematically recruited during L2 speaking, but explicit knowledge was preferentially used for correcting ungrammatical sentences in the EIT.

Although the results for the listening phase did not support the hypothesis, they nonetheless exhibited an intriguing pattern. In the listening phase, L2 speakers' metalinguistic knowledge was positively correlated with their left IFG activation (which was also activated by L1 speakers, indicating efficient access to procedural memory). In contrast, the metalinguistic score was negatively correlated with the left hippocampus (declarative memory) activity. Hence, it was explicit knowledge, rather than implicit knowledge, that influenced procedural memory retrieval, indicating a potential dynamic interaction between explicit and implicit knowledge. Presumably, L2 speakers with greater explicit knowledge preferentially used implicit knowledge during real-time grammar comprehension while inhibiting or minimizing access to explicit knowledge. This competing relationship between explicit and implicit knowledge systems is consistent with the theoretical predictions by Ullman's declarative-procedural model (Ullman, 2020), which was partially supported by an ad-hoc correlation analysis between left LIFG and left hippocampus showing a negative weak correlation ($r = -0.32$, $p = .12$).

In summary, the neural evidence presented here corroborates the aforementioned behavioral results. It further reveals that both explicit and implicit knowledge are deployed for the EIT and interact dynamically during the listening and speaking processes.

Limitations and suggestions for future research

Based on the current findings, as well as the inherent study limitations, we offer three major suggestions for future research directions. First, our sample size was small due to the labor- and cost-intensive nature of fMRI experiments. Thus, as brain imaging techniques allow us to examine the neurocognitive processes directly, conceptual replications of this experiment with different L2 learners would be highly valuable. Second, a retrospective questionnaire is an overly coarse instrument to measure awareness. It is possible that the ability to notice errors *and* provide examples of sentences containing errors (the criterion for higher-level awareness) is influenced by other factors such as individuals' memory ability and metalinguistic knowledge. Therefore, we must exercise caution when interpreting data gathered through the retrospective questionnaire. Third, the instructions of the EIT were not without problems.

It is possible that more indirect instructions used in previous research could have been ideal to lower the level of awareness of grammatical errors. In Erlam and colleague's research (e.g., Erlam, 2006; Erlam & Wei, 2021), for instance, participants were told to repeat sentences in correct English, and as part of the practice session before the main EIT, eight stimulus sentenes (four grammatical and four ungrammatical), were provided auditorily to show how ungrammatical sentences were converted to grammatical sentences. As their participants were not explicitly told that they should make such corrections, the instructions they received could be considered more "implicit" than those provided in the current study. Their instructions, however, may raise concerns about the reliability of the EIT. If some participants cannot notice errors in the instructions phase and/or fail to understand that they need to correct errors embedded in the stimuli, interpreting their EIT score becomes challenging, as their verbatim response without error correction may be intentional. Therefore, we ensured that all participants understood the written instructions accurately and discouraged them from repeating ungrammatical sentences verbatim. Nonetheless, the instructions in the present study to correct ungrammatical sentences might have pushed the L2 speakers to rely on explicit knowledge more than in the previous studies (e.g., Erlam, 2006; Erlam & Wei, 2021). To avoid this potential methodological limitation of EITs featuring ungrammatical sentences, some researchers have started using grammatical items only (Sarandi, 2020).

Conclusions: theoretical and methodological implications

As measurement is a scientific foundation of SLA theory, an ample body of research has been dedicated to the development of precise measures of L2 implicit knowledge. Because extant findings provide behavioral evidence both in favor of (e.g., Erlam, 2006; R. Ellis, 2009) and against (Spada et al., 2015; Suzuki & DeKeyser, 2015) EIT as a measure of implicit knowledge, the aim of this fMRI study was to reconcile these differences. Our findings demonstrated that L2 speakers dynamically use implicit knowledge as well as explicit knowledge to perform the listening and speaking EIT components, and their methodological and theoretical implications are discussed below.

The behavioral and neural evidence presented here suggests that advanced L2 speakers are skillful at using explicit knowledge for production during an EIT even under time pressure. While L2 speakers largely rely on explicit knowledge for EIT performance, proficient learners with some L2 experience may also draw upon implicit knowledge for certain skills such as comprehension, if not in all linguistic domains (DeKeyser, 2017; Paradis, 2009). Thus, it seems unwarranted to claim that L2 production tasks such as EIT can be a pure measure of solely speeded-up explicit *or* implicit knowledge, as the EIT scores obtained by most L2 speakers will be influenced by their explicit *and* implicit knowledge use. A broader implication of this assertion may be that EIT score is a strong indicator of general proficiency, because both implicit and explicit knowledge need to be integrated for fluent comprehension and production. In that case, to what extent it is useful to (attempt to) isolate implicit from explicit knowledge in production tasks is an important theoretical and fundamental question.

If L2 users can manage to speak accurately and fluently by efficiently retrieving explicit knowledge with minimal use of implicit knowledge, there may be trivial practical value of prioritizing the assessment of implicit over explicit knowledge. Although lack of awareness is the hallmark of implicit knowledge, awareness may not serve as a useful criterion to assess linguistic knowledge that is highly functional for L2 comprehension and production (e.g., DeKeyser, 2003; Paradis, 2009; Spada, 2015; Suzuki, 2017). Hence, we propose automaticity as an overarching feature of functional language processing and skills that are enabled by both implicit and speeded-up explicit knowledge use (DeKeyser, 2003, 2017; N. C. Ellis, 2015; R. Ellis, 2009). As automatic language processing is fast and efficient, and thus requires minimal mental resources, automatization of L2 knowledge and processing (regardless of awareness) is crucial to meet the processing demand of listening and time pressure imposed during speech process in the EIT. Indeed, some L2 researchers have already started to conceptualize EIT as a measure of automaticity on the premise that EIT can capture "procedural oral language ability and *degree of automaticity* in their speech" (Albarqi & Tavakoli, 2022, p. 7, emphasis added) and "implicit, as well as *automatized processing*" (Murakami & Ellis, 2022, p. 7, emphasis added). While acknowledging that distinct systems are involved in explicit and implicit learning (DeKeyser, 2003), the outcome of such learning may be more parsimoniously characterized by a single yet multifaceted construct—automaticity—that reflects efficient use of explicit and implicit knowledge for fluent comprehension and production skills.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by Joint Research Program of Joint Usage/Research Center at the Institute of Development, Aging, and Cancer at Tohoku University and Grant–in–Aid for Scientific Research (KAKENHI) from Japan Society for the Promotion of Science for the first and the second authors (15K02745, 18K00776, 21H00550). We are grateful to Masatoshi Sugiura for his constructive feedback on our earlier draft. We would also like to show our gratitude to the editors of this special issue, Megan Solon and Lauren Park, and anonymous reviewers for their helpful suggestions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.rmal.2023.100051.

References

Albarqi, G., & Tavakoli, P. (2022). The effects of proficiency level and dual-task condition on L2 self-monitoring behavior. *Studies in Second Language Acquisition*, 1–22. 10.1017/S0272263122000146.

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 33*(2), 247–271. 10.1017/S0272263110000756.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using the marsbar toolbox for SPM 99. *Neuroimage, 16*(2), S497. 10.1016/S1053-8119(02)90013-3.

Chrabaszcz, A., & Jiang, N. (2014). The role of the native language in the use of the English nongeneric definite article by L2 learners: A cross-linguistic comparison. *Second Language Research, 30*(3), 351–379. 10.1177/0267658313493432.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty, & H. M. Long (Eds.), *The handbook of second language acquisition* (pp. 312–348). Oxford: Blackwell Publishers.

DeKeyser, R. M. (2017). Knowledge and skill in ISLA. In S. Loewen, & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 15–32). New York, NY: Routledge.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al., (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage, 25*(4), 1325–1335. 10.1016/j.neuroimage.2004.12.034.

Ellis, N. C. (2015). Implicit and explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 3–24). Amsterdam, the Netherlands: John Benjamins.

Ellis, R. (2009). Measuring implicit and explicit knolwedge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31–64). Tonawanda, NY: Multilingual Matters.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*(3), 464–491. 10.1093/applin/aml001.

Erlam, R., & Wei, L. (2021). The importance of increased processing demands in the design of elicited imitation tests. *Language Teaching Research*, Article 13621688211026032. 10.1177/13621688211026032.

Godfroid, A., & Kim, K. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition, 43*(3), 606–634. 10.1017/S0272263121000085.

Granena, G. (2016). Elicited imitation as a measure of implicit L2 knowledge: The role of working memory and short-term memory capacity. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 185–204). Amsterdam: John Benjamins.

Isbell, D. R., & Rogers, J. (2021). Measuring implicit and explicit learning and knowledge. In P. Winke, & T. Brunfaut (Eds.), *The routledge handbook of second language acquisition and language testing* (pp. 304–313). New York: Routledge.

Jiang, N. (2011). *Conducting reaction time research in second language studies*. New York, NY: Routledge.

Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency: A meta-analysis. *Studies in Second Language Acquisition, 44*, 886–911. 10.1017/S0272263121000395.

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

Mizumoto, A. (2022). Calculating the relative importance of multiple regression predictor variables using dominance analysis and random forests. *Language Learning, Early View*. 10.1111/lang.12518.

Morgan-Short, K., Deng, Z., Brill-Schuetz, K. A., Faretta-Stutenberg, M., Wong, P. C. M., & Wong, F. C. K. (2015). A view of the neural representation of second language syntax through artificial language learning under implicit context of exposure. *Studies in Second Language Acquisition, 37*(s2), 383–419. 10.1017/S0272263115000030.

Murakami, A., & Ellis, N. C. (2022). Effects of availability, contingency, and formulaicity on the accuracy of English grammatical morphemes in second language writing. *Language Learning, Early View*. 10.1111/lang.12500.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Philadelphia, PA: John Benjamins Publishing Company.

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 878–912. 10.1111/lang.12079.

Roehr-Brackin, K. (2022). Explicit and implicit knowledge and learning of an additional language: A research agenda. *Language Teaching*, 1–19. 10.1017/S026144482200026X.

Sarandi, H. (2020). The structural sensitivity of elicited imitation as a measure of implicit grammatical knowledge. *Eurasian Journal of Applied Linguistics, 6*(2), 265–284. 10.32601/ejal.775806.

Seok, J.-. W., & Cheong, C. (2020). Functional dissociation of hippocampal subregions corresponding to memory types and stages. *Journal of Physiological Anthropology, 39*(1), 15. 10.1186/s40101-020-00225-x.

Spada, N. (2015). SLA research and L2 pedagogy: Misapplications and questions of relevance. *Language Teaching, 48*(1), 69–81. 10.1017/S026144481200050X.

Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning, 65*(3), 723–751. 10.1111/lang.12129.

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics, 38*(5), 1229–1261. 10.1017/S014271641700011X.

Suzuki, Y., & DeKeyser, R. M. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning, 65*(4), 860–895. 10.1111/lang.12138.

Suzuki, Y. (2022). Automatization and practice. In A. Godfroid, & H. Hopp (Eds.), *The Routledge handbook of second language acquisition and psycholinguistics* (pp. 308–321). New York: Routledge.

Suzuki, Y., Jeong, H., Cui, H., Okamoto, K., Kawashima, R., & Sugiura, M. (2022). An fMRI validation study of the word-monitoring task as a measure of implicit knowledge: Exploring the role of explicit and implicit aptitudes in behavioral and neural processing. *Studies in Second Language Acquisition, Early View*. 10.1017/S0272263122000043.

Tagarelli, K. M., Shattuck, K. F., Turkeltaub, P. E., & Ullman, M. T. (2019). Language learning in the adult brain: A neuroanatomical meta-analysis of lexical and grammatical learning. *Neuroimage, 193*, 178–200. 10.1016/j.neuroimage.2019.02.061.

Ullman, M. T. (2020). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten, G. D. Keating, & S. Wulff (Eds.), *Theories in second language acquisition: An introduction* (3rd ed.) (pp. 128–161). New York, NY: Routledge.

Vafaee, P., Suzuki, Y., & Kachinske, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition, 39*(1), 59–95. 10.1017/S0272263115000455.

Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing, 33*(4), 497–528. 10.1177/0265532215594643.

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition, 37*(3), 457–486. 10.1017/S0272263114000370.