Studies in Second Language Acquisition (2025), **47**, 1018–1043
doi:10.1017/S0272263125101290

RESEARCH ARTICLE

Timing matters for interactive task-based learning

Effects of vocabulary practice on learning multiword expressions and neural synchronization

Yuichi Suzuki[1], Takayuki Nozawa[2], Takumi Uchihara[3], Sachiko Nakamura[4], Atsuko Miyazaki[5] and Hyeonjeong Jeong[3]

[1]Faculty of International Research and Education, Waseda University, Tokyo, Japan; [2]Faculty of Engineering, University of Toyama, Toyama, Japan; [3]Graduate School of International Cultural Studies, Tohoku University, Sendai, Japan; [4]Center for English as a Lingua Franca, Tamagawa University, Tokyo, Japan and [5]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan
Corresponding author: Yuichi Suzuki; Email: yszk@waseda.jp

(Received 05 February 2025; Revised 13 July 2025; Accepted 11 August 2025)

Abstract

To investigate the effects of vocabulary practice timing on lexical learning and neural processing during communication tasks, we engaged 80 Japanese learners of English (40 pairs) in information-gap tasks with vocabulary practice in the pre-task or post-task phase. Learning of verb-noun combinations was orally assessed through translation and narrative tasks one week later. To quantify neural coupling between participants during task-based interaction, hyperscanning with fNIRS was used to measure inter-brain synchronization (IBS). Results showed that pre-task vocabulary practice led to greater learning, while post-task practice resulted in higher IBS in the brain region underlying language processing. Better vocabulary practice performance predicted more automatic use of multiword expressions in the post-task condition. IBS in the brain regions underlying social cognition and language processing predicted multiword learning. These findings reveal how practice timing influences neural synchronization and language acquisition, highlighting the importance of balancing lexical learning with communication processes in task-based language teaching.

Keywords: Instructed SLA; Inter-brain synchronization; Multiword expressions; Task-based language Teaching; Vocabulary practice

Introduction

Task-based language teaching (TBLT) has gained prominence in second language acquisition (SLA) research, emphasizing the benefits of "learning by doing" through meaningful communication (R. Ellis et al., 2020; Long, 2015; Samuda & Bygate, 2008; Skehan, 1998). This approach aligns with the usage-based perspective on language learning, highlighting the role of social interaction in knowledge acquisition (N. C. Ellis & Wulff, 2015; Tyler & Ortega, 2018).

While TBLT primarily relies on incidental learning—where learning occurs as a by-product of engaging in communicative tasks (Newton, 2013), it also raises questions about integrating form-focused instruction within communicative L2 use (R. Ellis et al., 2020). Vocabulary scholars have explored various approaches to enhance incidental learning through reading (e.g., Laufer & Girsai, 2008; Yanagisawa et al., 2020) and listening (e.g., Zhang & Graham, 2020) tasks, but TBLT's application to vocabulary instruction remains less explored compared to grammar (e.g., Fotos & Ellis, 1991) and pronunciation (Mora & Levkina, 2017).

Likewise, the timing of form-focused instruction has been extensively investigated in contextualized reading and writing tasks (Altamimi & Conklin, 2024; Elgort et al., 2024; Pellicer-Sánchez et al., 2021; Webb, 2009), whereas the optimal timing for vocabulary instruction integration in speaking tasks remains unexplored. This is a key issue in TBLT, as evident from scholarly support for both pre-task (Willis & Willis, 2007) and post-task (R. Ellis, 2003; Skehan, 1998) lexical instruction.

The complexity of speaking presents unique challenges for vocabulary acquisition in L2 communication, necessitating consideration of multiple dimensions of lexical competence (Cheng et al., 2023; Saito et al., 2025; Uchihara et al., 2025). Unlike receptive skills such as reading and listening, speaking requires learners to simultaneously orchestrate conceptualization, linguistic formulation, and articulation processes (Levelt, 1989). While writing allows for extended pausing and revision, speaking requires quick lexical retrieval; L2 learners often face cognitive demands that limit attentional resources available for using appropriate lexical items in context in real time.

In the TBLT context, the role of formulaic language for successful communication deserves particular attention, as formulaic language, such as collocations and idioms, plays a crucial role in fluent language production (e.g., Siyanova-Chanturia & Pellicer-Sánchez, 2018; Wray, 2002). Using multiword expressions (i.e., combinations of individual words such as verb + noun and adjective + noun) contributes significantly to conveying meaning essential for task completion, potentially making them more valuable than single lexical items for successful communication in speaking tasks (Takizawa, 2024).

Furthermore, recent advances in neuroscience have introduced hyperscanning as a promising technique for investigating L2 learning processes during communication. This neuroimaging method enables simultaneous measurement of brain activity in multiple interacting individuals, providing an index of inter-brain synchronization (IBS) during communication tasks (for recent reviews, see Kelsen et al., 2022; Zhang et al., 2024). The temporal coordination of neural activity between individuals engaged in social interaction, as measured by IBS, offers new insights into the neural processes underlying L2 learning through communication and the effects of instructional interventions. Specifically, the timing of vocabulary practice presumably influences how learners allocate attention during interaction, thereby potentially modulating the degree of neural coupling (i.e., IBS) between partners.

This study aims to extend our understanding of how the timing of vocabulary practice influences both neural processes and lexical learning through interactive information-gap tasks. Specifically, pre-task practice may alter how learners allocate attention, thus affecting the quality of interpersonal coordination and neural coupling (IBS) between them. We thus examine how this timing difference impacts both lexical learning and the underlying neural synchronization during interaction. By integrating perspectives from TBLT, vocabulary acquisition research, and cognitive neuroscience, we seek to provide unique insights into optimizing vocabulary learning in communicative contexts.

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

Background literature

Timing of lexical instruction in task-based language learning

Research on L2 vocabulary acquisition has demonstrated the effectiveness of explicit pre-task lexical instruction for the acquisition of both single and multiword lexical items in reading and writing tasks (Altamimi & Conklin, 2024; Pellicer-Sánchez et al., 2021; Webb, 2009). Although the obtained results mostly demonstrate that pre-teaching can contribute to the development of vocabulary knowledge measured through written modalities in reading and writing tasks, its effectiveness in speaking instruction remains unclear. Thus, investigating speaking tasks presents an opportunity to extend research on lexical learning through interactive information-gap tasks.

Existing evidence indicates that vocabulary can be learned through interaction and negotiation of meaning (e.g., R. Ellis & He, 1999; Newton, 2013). To further promote vocabulary learning through meaning-focused output using communication tasks, integrating meaning-focused output with language-focused practice is critical from the perspectives of both vocabulary acquisition and TBLT research (e.g., Suzuki, 2023). According to Nation's (2007) four-strand principle, for instance, meaning-focused output and language-focused learning need to be balanced along with meaning-focused input and fluency development.

However, to facilitate vocabulary acquisition through social interaction, it is essential to establish *when* language-focused practice can be optimally integrated into the communication tasks. The optimal timing of vocabulary support (e.g., pre-task versus post-task) must balance the demands of oral interaction with successful task completion. While some TBLT scholars acknowledge the potential benefits of pre-task lexical instruction (Willis & Willis, 2007), others find them problematic, as learners might treat the task as an exercise of target vocabulary items (R. Ellis, 2003; Skehan, 1998).

Thus far, such debates on the optimal integration of lexical instruction have largely remained theoretical. As one of the exceptions, de la Fuente (2006) compared the effectiveness of pre-task and post-task vocabulary practice for lexical acquisition through information-gap and role-play tasks. While both conditions led to similar performance in immediate retrieval of target words, the post-task group demonstrated significantly better results in the delayed posttest conducted one week later. However, the study included several procedural differences between groups beyond lexical instructional timing, such as task planning and reporting with incidental focus-on-form. These additional variables make it difficult to isolate the effects of vocabulary practice timing alone.

Accordingly, in the present study, the effects of pre-task and post-task vocabulary practice on the acquisition of multiword expressions through information-gap tasks are compared. The aim is to elucidate how the benefits of social learning contexts can be effectively combined with targeted language-focused practice to optimize vocabulary acquisition in speaking tasks.

The role of vocabulary practice in communication task implementation

In TBLT, the role of deliberate vocabulary practice requires careful consideration. On the one hand, it is important to prepare learners to actively use target vocabulary in meaning-focused output activities by providing sufficient pre-task vocabulary practice

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

to develop productive lexical knowledge (Schmitt, 2008). Consequently, a pre-task vocabulary practice session is often included in reading studies and learners are trained until they memorize all target words (e.g., Pellicer-Sánchez et al., 2021). On the other hand, from a TBLT perspective, such extensive pre-task deliberate vocabulary instruction may transform communication tasks into mere language exercises, which is not the intended outcome (e.g., R. Ellis, 2003; Skehan, 1998).

To address this challenge, deliberate practice in the pre-task stage should be carefully designed (e.g., R. Ellis et al., 2019; R. Ellis et al., 2020). It should provide enough support to facilitate task performance without being so extensive as to undermine the communicative nature of the task. This approach allows for individual variations in vocabulary knowledge among learners while preserving the core principles of task-based learning.

While the role of pre-task vocabulary practice has received considerable attention in TBLT literature, empirical investigations of post-task vocabulary practice remain limited. The aforementioned study by De la Fuente's (2006) provides some preliminary evidence suggesting potential benefits of post-task lexical instruction, possibly because learners first activate conceptual content through the communication task and can then connect this contextual information to lexical forms practiced afterward. The scarcity of systematic comparisons highlights a clear need for research that directly investigates the effects of practice timing.

Guided by this premise, the current study investigates how the performance in a brief deliberate practice during the pre-task stage may influence the subsequent vocabulary acquisition through communication tasks. The findings may elucidate when vocabulary practice is most effective, as well as how its quality (as measured by vocabulary practice performance) influences lexical learning in communication tasks. Possibly, more robust form–meaning mapping knowledge could lead to the development of lexical knowledge in communicative contexts (Schmitt, 2008).

The multifaceted nature of vocabulary knowledge for L2 speaking skills

Recognizing the importance of integrating language-focused practice with social learning contexts, it is crucial to consider the multifaceted nature of vocabulary knowledge, particularly in relation to L2 speaking skills. Whereas traditional vocabulary research tends to focus primarily on declarative knowledge, Nation's (2022) influential componential model emphasizes both form–meaning connections and contextual use. Psycholinguistic approaches have expanded this framework (Godfroid, 2019; Uchihara, 2025), identifying four key features of multifaceted vocabulary knowledge. First, dimension (recall vs. recognition) refers to the knowledge strength continuum from recognition to recall, with recall requiring stronger form-meaning connections essential for speech production (González-fernández & Schmitt, 2020). Second, formulaicity (single vs. multiword items) encompasses knowledge of multiword expressions that constitute a significant portion of natural discourse and contribute to fluent production (Siyanova-Chanturia & Pellicer-Sánchez, 2018). Third, automaticity (online vs. offline) is a complex construct beyond mere speed, referring to efficient and stable lexical access, crucial for real-time speaking demands (DeKeyser, 2001; Segalowitz, 2010; Suzuki et al., 2025). Fourth, context (contextualized vs. decontextualized) distinguishes between isolated word knowledge and appropriate usage within communicative contexts (Nation, 2022; Read, 2000).

While this multifaceted view of lexical knowledge has informed an emerging line of research on receptive skills such as listening (Saito et al., 2025; Uchihara et al., 2025), L2

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

speaking involves distinct active processes of conceptualization, formulation, and articulation (Levelt, 1989), necessitating specific adaptations for assessment. We thus extend the measurement framework established by Saito et al. (2025) and Uchihara et al. (2025) for measuring L2 word knowledge employable for real-life listening—which emphasize the assessment of both declarative and procedural aspects of vocabulary knowledge in both contextualized and decontextualized settings, this study focuses on two aspects of productive lexical knowledge (Read, 2000): *recall* and *use* of the spoken forms of word combinations. Recall of L2 forms can underlie controlled and decontextualized knowledge of form−meaning mapping (measured via a translation task), whereas their use can be characterized as spontaneous and contextualized L2 knowledge (elicited through a narrative task). Accuracy and speed measures can be derived from the translation and narrative tasks to assess declarative knowledge and processing automaticity of verb-noun combinations, respectively (Godfroid, 2019; Saito et al., 2025; Uchihara et al., 2025). This approach allows the impact of pre-task vocabulary practice on both the declarative knowledge and processing automaticity needed for communicative contexts to be examined.

The role of IBS in L2 communication and vocabulary learning

Recent advances in neuroscience have provided new ways of evaluating the role of social interaction in L2 acquisition. L2 acquisition through collaborative problem-solving and joint attention through task-based interaction is promoted in both TBLT (e.g., Long, 2015; R. Ellis et al., 2020) and usage-based approach to language learning (e.g., N. C. Ellis & Wulff, 2015; Tyler & Ortega, 2018).

Neuroscientific evidence supports this theoretical stance, demonstrating that social context engages a broader network of brain regions associated with action perception and social-affective processes during language learning, compared to decontextualized methods like vocabulary learning through translation (see Li & Jeong, 2020, for review). Building on these findings, neuroscientists have recognized the need to study brain activity in more naturalistic, interactive settings. This realization has led to the emergence of "second-person neuroscience," a new approach that shifts the focus from studying individuals in isolation to examining brain activity during real-time social encounters (e.g., Schilbach et al., 2013).

In this context, hyperscanning is particularly valuable, as it enables simultaneous measurement of brain activity in two or more interacting individuals. Functional near-infrared spectroscopy (fNIRS) is advantageous for such research as it has fewer constraints compared to methods like functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), making it well-suited for investigations in real-world settings. fNIRS differs from fMRI and EEG regarding spatial resolution, temporal resolution, and tolerance for participant movement (for an overview, see Quaresima & Ferrari, 2019). Specifically, fNIRS offers better spatial resolution than EEG, advantageous for reliably capturing cortical activity (e.g., in the prefrontal cortex) during interaction. Although EEG provides superior temporal resolution by directly measuring electrical activity, EEG is highly sensitive to motion and muscle artifacts. Compared to fMRI, fNIRS typically has a higher temporal resolution. Crucially for studying dynamic social interactions, fNIRS provides greater tolerance for participant movement and less restraint than EEG or fMRI, thus allowing studies in more naturalistic settings where participants can comfortably sit and converse (see Figure 2 for an experimental setup).

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

This flexibility allows researchers to measure changes in brain hemodynamics and infer neural activity during social interactions (e.g., Zhang et al., 2024). In particular, in fNIRS hyperscanning studies, inter-brain synchronization (IBS) has been observed in frontal and temporoparietal brain regions during cooperative tasks requiring shared attention, mentalizing, and collaborative problem-solving (Kelsen et al., 2022). Neuroscience research indicates functional specialization within the prefrontal cortex (PFC), with lateral PFC (LPFC) often linked to executive functions including language processing, and medial PFC (mPFC) more involved in social cognition such as mentalizing (Gilbert et al., 2006). A growing body of neuroscience research has demonstrated that higher IBS levels are associated with enhanced learning outcomes across various educational contexts (Zhang et al., 2024). Specifically, stronger neural coupling, indexed by IBS in both lateral and medial frontal regions, as well as temporoparietal regions, correlates with better performance in collaborative tasks such as discussion, video gaming, and singing (Tan et al., 2023). Therefore, while IBS in both left (LPFC) and mPFC has been associated with enhanced learning outcomes, the specific brain regions whose synchronization is most critical for learning may vary depending on the nature of the task and the learning aspects being measured. Application of hyperscanning techniques to Instructed SLA research is thus expected to extend the current knowledge of the neural processes underlying task-based communication (e.g., Gurzynski-Weiss, 2024).

In a pioneering fNIRS hyperscanning study, Nozawa et al. (2019) found that prior physical synchrony between paired L2 learners enhanced IBS and rapport in the left LPFC during language-focused vocabulary practice (i.e., teaching unfamiliar English words in learner dyads). This finding suggests that neural synchrony during paired vocabulary practice may contribute to enhanced vocabulary knowledge acquisition. However, this investigation focused on decontextualized (e.g., word lists) rather than contextualized (through communication tasks) vocabulary learning, leaving the effects of vocabulary practice on IBS during interactive tasks unexplored.

To our knowledge, no research has investigated the effects of pre-task vocabulary teaching on IBS during communication tasks. While some TBLT researchers promote pre-task vocabulary teaching prior to the main task (Willis & Willis, 2007), pre-task vocabulary practice could potentially disrupt communication and divert attention from the task itself (e.g., R. Ellis, 2003; R. Ellis et al., 2019; Skehan, 1998). These divergent perspectives raise the question of whether such pre-task practice might increase or decrease IBS when a communication task is preceded by vocabulary practice, as L2 learners' attention is divided to focus more on producing the practiced vocabulary items accurately and to coordinate interaction for mutual understanding with their partners.

By examining IBS in communication tasks, researchers may uncover the neural mechanisms underlying social learning processes. Given the importance of social interaction in language learning, investigating how IBS relates to lexical learning could provide valuable information about both the neurocognitive "process" during tasks and the "product" of lexical acquisition. In sum, four significant gaps identified in this literature review are addressed in the current study:

1. The relative effects of pre-task and post-task vocabulary practice on lexical learning
2. The relationship between vocabulary practice performance and subsequent lexical learning
3. The effects of pre-task vocabulary practice on IBS during interactive communication tasks
4. The potential relationships between IBS and lexical learning

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

This integration of neuroscientific methods with instructed SLA research approaches (e.g., Gurzynski-Weiss, 2024) offers a promising avenue for advancing our understanding of L2 learning processes in interactive contexts.

The current study

This study investigated the effects of vocabulary practice timing on lexical learning and IBS during interactive communication tasks. We employed a pretest–posttest design with random assignment to experimental conditions (pre-task or post-task vocabulary practice). Participants were 80 Japanese learners of English as an L2 (40 dyads), who engaged in information-gap (communication) tasks. In the pre-task vocabulary practice condition, participants practiced target verb-noun combinations before the communication tasks, while in the post-task vocabulary practice condition, they practiced the same verb-noun combinations after completing the communication tasks. During both vocabulary practice and communication tasks, participants' medial prefrontal neural activity was measured using fNIRS to assess IBS. Lexical learning was evaluated through a pretest and a one-week delayed posttest.

Four research questions (RQs) were addressed:

RQ1. To what extent does the timing of vocabulary practice influence lexical learning?
RQ2. To what extent does the vocabulary practice performance predict lexical learning?
RQ3. To what extent does pre-task vocabulary practice influence IBS in the brain regions associated with language processing (left LPFC) and social cognition (mPFC) during communication tasks?
RQ4. To what extent does IBS in the mPFC and left LPFC during vocabulary practice and communication tasks predict lexical learning?

Guided by previous research, we formulated the following hypotheses:

H1. The pre-task group would exhibit greater lexical knowledge gains than the post-task group.
H2. Better vocabulary practice performance would lead to greater learning.
H3. Pre-task vocabulary practice would (a) increase IBS in the left LPFC (language processing) and (b) decrease IBS in the mPFC (social cognition) during subsequent communication tasks. Alternatively, it might decrease IBS in both left LPFC and mPFC.
H4. IBS particularly in the left LPFC (language processing) during vocabulary practice and communication tasks would positively predict lexical learning.

These hypotheses draw upon theoretical considerations and empirical findings from TBLT and neuroscience research. H1 stems from the idea that pre-task practice may enable more active and effective lexical use during subsequent communication, enhancing knowledge consolidation (Schmitt, 2008). H2 is based on the principle that successful initial learning during practice provides a stronger foundation for later application and deeper acquisition (e.g., Nation, 2007). H3 considers both the functional specialization within the PFC (left LPFC for language processing, mPFC for social cognition; Gilbert et al., 2006) and the TBLT notion of a potential trade-off between attention to linguistic forms and communication (e.g., R. Ellis, 2003; R. Ellis

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

et al., 2020; Skehan, 1998). Integrating these, we primarily predict that pre-task practice, by orienting attention to linguistic forms, would exert differential influence on IBS in (a) the left LPFC (language processing) and (b) the mPFC (social cognition) during subsequent communication tasks. This functional brain specialization might lead to (a) increased IBS in the left LPFC and (b) decreased IBS in the mPFC due to reduced interpersonal focus. However, an alternative possibility also exists: if pre-task practice shifts learner's attention heavily towards lexical processing *within* themselves, it might reduce overall inter-personal coordination, potentially leading to decreased IBS in the left LPFC (hence, reduced IBS in both regions). The lack of prior fNIRS hyperscanning research in interactive L2 learning contexts hinders strong directional IBS predictions. Hence, the predictions for H3 should be considered exploratory. Finally, H4 builds on findings linking IBS to learning outcomes (e.g., Tan et al., 2023; Zhang et al., 2024) and posits that, in the context of lexical acquisition facilitated by practice, synchronization in the brain region most relevant to language processing (left LPFC) would be particularly predictive of learning gains.

Method

Participants

The study sample comprised 80 English learners (48 female, 32 male; age 18−22 years, M = 19.25, SD = 1.19) with Japanese as their first language (L1).[1] They were university students and their academic backgrounds spanned various disciplines within the humanities and social sciences, including cross-cultural studies, foreign languages, and international business. All participants were right-handed and had limited exposure to English-speaking environments, with no individual having spent more than three months studying or residing in an Anglophone country.

Given the interactive nature of the primary task, we controlled for potential confounding variables by recruiting same-sex pairs of already acquainted individuals. We collected data on both general language proficiency (elicited imitation task; Wu et al., 2022) and dyad interaction (face-to-face meetings and digital communication) frequency. Analyses revealed comparable proficiency levels between groups but significantly higher pair contact frequency in the post-task group (see Appendix A in Online Supplementary Material for details). To account for these participant characteristics, both proficiency scores and pair contact frequency were included as covariates in all subsequent analyses.

Communication tasks

Two interactive information-gap tasks—"Company Spy" and "Museum Thief"—were adapted from a recently published TBLT textbook designed to facilitate English interactive skills development (Kelly & Suzuki, 2025). Both tasks required participants to work in dyads, assuming sender and receiver roles to exchange information and achieve a specific objective (see Figure 1).

[1]Initially, 84 university students (forming 42 same-sex pairs) were recruited for the study. However, due to unforeseen illness, one participant from each of two separate pairs was unable to attend the second experimental session. As the study design required pair participation across both sessions, data from these two pairs (four participants in total) were excluded from the final analysis.

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

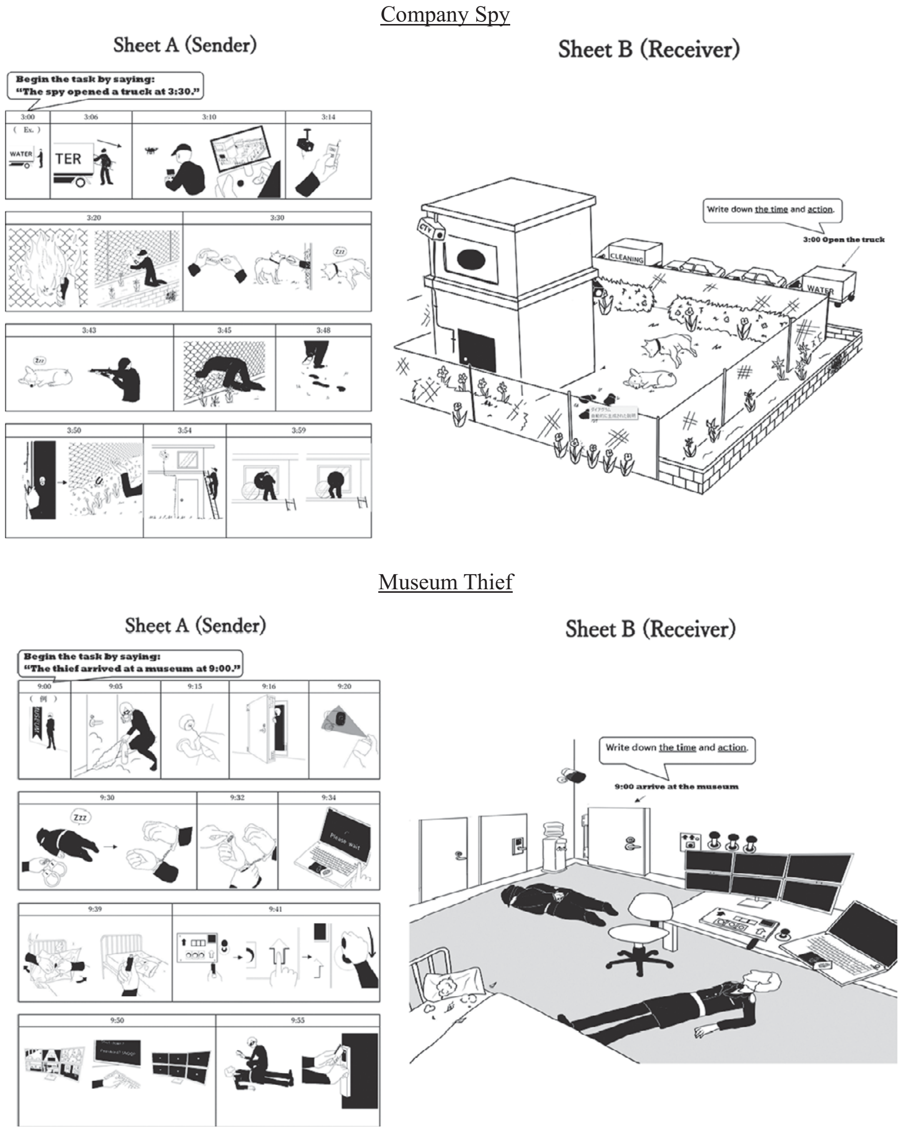

Figure 1. Communication tasks: Company Spy Task (top) and Museum Thief (bottom). *Note*: Enlarged materials are available in Appendix B in Online Supplementary Material.

In the "Company Spy" task, the primary goal was to reconstruct the spy's infiltration route through a company's premises. The sender was provided with a handout depicting security camera footage, which chronicled the spy's movements with corresponding time stamps. The receiver, equipped with a map of the building, was tasked with documenting the spy's progress based on the sender's descriptions. For example, the sender might communicate, "The spy climbed a ladder to the second floor at 9:35." Receivers were encouraged to seek clarification when necessary, promoting the interactive aspect of task-based communication.

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

The "Museum Thief" task followed a similar structure, with participants working to reconstruct a thief's path through a museum. As in the previous task, senders described the thief's actions and provided timestamps (e.g., "The thief released sleeping gas at 9:05") based on the provided video footage screenshots, while receivers documented this information on a museum floor plan. This parallel design ensured consistency across tasks while presenting a novel scenario to maintain participant engagement.

This task design aligns with the key TBLT research principles, emphasizing authentic communication, information exchange, and the negotiation of meaning within a structured context (R. Ellis et al., 2020). Both tasks featured an equal number of scenes (19) to describe and shared a similar narrative structure, requiring minimal causal reasoning.

Target vocabulary

We selected 22 verb–noun combinations (see Appendix C in Online Supplementary File) that are useful in effectively completing the communication tasks. Eleven items were extracted from each of the two tasks: Company Spy (e.g., "fly a drone," "take out a ladder," "shoot a gun") and Museum Thief (e.g., "release gas," "unlock a door," "turn on a computer").

All verbs and most nouns were among the most frequent 2000 word families in the BNC/COCA-25 lists (Nation, 2017). This suggests that while participants might recognize the individual words within the verb–noun combinations, they might still struggle to produce these specific combinations. This assumption was supported by classroom observations from prior implementations of the target information-gap tasks, where two of the authors had noted that Japanese university students did not always produce these specific combinations. These characteristics made the selected verb-noun combinations potential pedagogical targets that would benefit from deliberate vocabulary practice.

To ensure the naturalness and appropriateness of our target items, we consulted with a native English speaker who reviewed each verb-noun combination alongside the corresponding scene. All word combinations were verified as natural and appropriate expressions for describing the depicted actions in those specific contexts. It is important to note that we selected these verb–noun combinations based on their functional importance for our communication tasks rather than on statistical frequency criteria that is required in the definition of collocations (see Appendix C for the t-scores and MI scores for the verb + noun combinations). Some of the combinations are collocations (e.g., "fly a drone," "shoot a gun," "put handcuffs") according to statistical criteria (e.g., MI scores > 3; Hunston, 2002), while others (e.g., "put a drug," "burn a fence," "release gas," "tear up a pillow") do not meet these strict statistical thresholds. Our approach aligns with a research perspective emphasizing contextual relevance in vocabulary learning (Wood, 2019) and is justifiable from a practical perspective such as TBLT for preparing linguistic resources that are useful for task completion.

Vocabulary practice sheet

The vocabulary practice sheet contained a Japanese–English list of the target 22 verb–noun combinations (see Appendix D in Online Supplementary File). The word combinations were deliberately arranged in a different order from their appearance in the communication task narration to prevent learners from mechanically following the list during task completion. The sheet was used during the vocabulary practice session before or after the communication phase, where dyads took turns testing each other on the target multiword expressions (see the Procedure section for details).

fNIRS measurements

To assess prefrontal brain activity during face-to-face learning tasks (Figure 2a and b), we employed the HOT-2000 fNIRS device (NeU Corporation, Japan). This device utilizes near-infrared light to detect hemodynamic changes in the brain, which serve as neural activity indicators. The apparatus is designed as a comfortable headband with two sets of light emitters and detectors (Figure 2c), enabling natural interaction between participants while measuring activity in distinct PFC regions associated with external and internal mental processes.

We focused on the neural activity in the mPFC and left LPFC (see Figure 2d) during both vocabulary practice and communication task phases. This choice was guided by the distinct functional roles of these regions and previous findings of their synchronized activity during verbal communication. Specifically, the left LPFC is primarily implicated in language processing and production, while the mPFC is crucial for social cognition and mentalizing (Gilbert et al., 2006).

The fNIRS system quantifies total hemoglobin concentration changes, previously established as a reliable proxy for localized brain activity. To ensure optimal signal quality, we positioned the sensors on participants' foreheads, carefully displacing hair and employing light-blocking covers to minimize external interference. Participants were instructed to maintain a relaxed posture and minimize head movements to

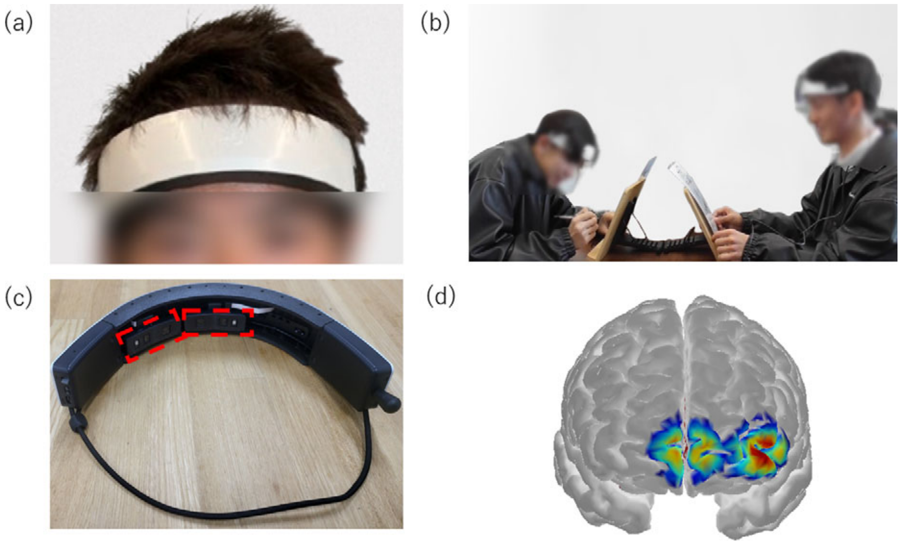

Figure 2. fNIRS measurement and learning task setup.
Note: Figure (d) shows the spatial sensitivity profile of the two fNIRS channels to cortical hemodynamic changes, generated using the HOMER and AtlasViewer software. Warmer colors indicate higher sensitivity to hemodynamic changes in these brain regions.

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

Table 1. Spoken vocabulary knowledge

	Accuracy	Speech rate
Translation task	Declarative knowledge of form–meaning mapping in translation	Processing efficiency for retrieving form–meaning mapping in translation
Narrative task	Declarative knowledge of lexical use in context	Processing efficiency for lexical use in context

enhance signal clarity and reduce motion-induced artifacts. While fNIRS is generally more tolerant to such artifacts than techniques like EEG (Quaresima & Ferrari, 2019), minimizing movement remains important for optimizing data quality.

Vocabulary learning measures

To assess spoken vocabulary knowledge, we employed translation and narrative tasks that measure different aspects of verb-noun combination knowledge (see Table 1). Building on the framework of multifaceted vocabulary knowledge as delineated in the Background Literature section (e.g., Godfroid, 2019; Read, 2000; Saito et al., 2025; Uchihara et al., 2025), our assessment distinguishes between accuracy (declarative knowledge) and speed (processing efficiency) in two contexts: (a) decontextualized recall and (b) contextualized use.

The translation task assesses the ability to recall the spoken forms of verb-noun combinations in a decontextualized situation when prompted with L1 translations. This evaluates both declarative knowledge of form–meaning mapping (accuracy of production) and processing efficiency for the retrieval of this mapping (speech rate). The narrative task focuses on the ability to use these combinations in context, evaluating declarative knowledge of lexical use (accuracy) as well as processing efficiency for lexical use (speech rate) within a meaningful communicative situation.

As established in recent research (Saito et al., 2025; Uchihara et al., 2025), these complementary tasks capture different aspects of vocabulary knowledge. The translation task primarily reflects form-meaning connections, while the narrative task additionally captures use-in-context knowledge. Processing efficiency is typically considered as an index of automaticity (DeKeyser, 2001; Suzuki et al., 2025); however, speech measures such as speech rate should not be equated with full automaticity, which typically requires restructuring of lexical knowledge representation (Segalowitz, 2010). While the exact nature of this restructuring process in vocabulary learning warrants further research, it presumably involves (a) strengthening form–meaning connections and (b) developing appropriate contextual use of lexical items. We do not claim that speech rate alone represents automatized lexical knowledge, but rather view fast performance (i.e., faster speech rate) as a meaningful product of learning that reflects degrees of automaticity, acknowledging that automaticity exists on a continuum from slow, hesitant L2 processing to efficient, fluent performance (DeKeyser, 2017; Suzuki, 2022).

Together, these tasks provide a comprehensive evaluation of learners' spoken vocabulary knowledge that encompasses both declarative knowledge and the degree to which this knowledge has become efficient for fluent use in communication.[2]

[2]Correlations among the accuracy and speed measures on the two tasks were weak to moderate ($.29 < r < .67$), suggesting that each measure taps into different aspects of vocabulary knowledge (see Appendix E in Online Supplementary File).

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

Figure 3. Translation task procedure.

Translation task

A translation (cued recall) task was employed to assess participants' lexical knowledge. As shown in Figure 3, participants viewed Japanese phrases on a computer screen and were asked to produce the corresponding English verb–noun combinations. For each English word, the first letter of each word was provided as a cue, with the number of remaining letters indicated by underscores (e.g., (f_ _) (a) (d _ _ _ _) for "fly a drone"). Participants were given 8 seconds to respond to each item and were instructed to respond as accurately and quickly as possible.

Responses were scored manually. Correct answers were awarded one point, based on pronunciation comprehensibility (Uchihara, 2022). Comprehensible responses, despite minor pronunciation and/or morphological errors that won't severely influence comprehensibility (e.g., "stole money" instead of "steal money," "leave footprint" instead of "leave footprints," "collect a fingerprints" instead of "collect a fingerprint"), received one point.

Speech duration was measured from the onset of the cue (i.e., Japanese translation) to the end of utterance for all fully correct items. To standardize utterance length across multiword expressions, speech rate (i.e., the number of syllables per second) was computed by dividing the number of syllables in multiword expressions by the speech duration for each item.

Two Japanese speakers initially coded responses from 11 participants for both accuracy and speed measures. Based on 242 responses in each measure, 99% intercoder agreement was achieved, and the remaining discrepancies were resolved through discussion. After achieving 100% agreement, one coder proceeded to score responses from the remaining participants. Satisfactory reliability for accuracy measures, indexed by Cronbach's alpha, was achieved for pretest (α = .67) and posttest (α = .78).

Narrative task

A narrative (monologue speech) task was used to measure lexical knowledge deployable in context. Participants were asked to orally narrate specific scenes without a time limit, allowing for the elicitation of lexical usage in a context. They narrated only the task for which they had acted as the information sender during the main communication activity. Thus, each participant received a handout for either the Company Spy or Museum Thief task, showing security camera footage of a spy's or thief's entry, respectively. Because we were interested in the effects of output during communication task on the lexical use in context, this approach ensured that participants' narratives on

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

this task were presumably based on their actual experience of describing the scenes during the interactive task.

Narrations were recorded and transcribed. Then, all the utterances with target multiword expressions were extracted from the transcripts and were coded for accuracy and speed measures. Consistent with the translation task scoring rubrics, one point was given for correct use of target items, disregarding minor morphological errors such as tense, number, or article mistakes. No credit was given when a participant did not use the target multiword expression when describing the relevant scene in the sequence, for instance, by using a different, nontarget expression or by describing the scene without the target phrase.

Speech duration was measured from the offset of words immediately preceding the target expression to the end of the target expression for correct productions, including any pauses that occurred before or within the expression. For instance, in the utterance "The spy flew a drone at 7 PM," the speech duration included both the speaking time and any pauses from the offset of "spy" to the end of "drone." This inclusive measurement of pauses provides a more complete picture of the temporal aspects of lexical production in spontaneous speech (see Suzuki, 2017 for a similar approach in grammar). In line with the narrative task, speech rate was computed by dividing the number of syllables in multiword expressions by the total speech duration (including pauses) for the multiword expressions.

Procedure

Participants were recruited at the university through flyer advertisements and were compensated for their participation in two sessions. Upon arrival in an office for the first session, all participants signed a consent form and were then randomly assigned to either the pre-task or the post-task group.

Using a pretest–posttest design, the first session consisted of pretest and two communication tasks along with vocabulary practice, and a delayed posttest was administered during the second session one week later, as illustrated in Figure 4.[3]

The first session commenced with the translation task administered as a pretest. Participants were then fitted with fNIRS devices, which they wore during the vocabulary practice and communication tasks. In the pre-task group, learners received a vocabulary practice sheet with a Japanese–English list of verb–noun combinations relevant to the communication tasks prior to engaging in them. To familiarize themselves with the target multiword expressions, participants listened via a speaker to an audio recording, which was generated specifically for this study using AI software with a standard female American English voice. Subsequently, the activity transitioned into a paired vocabulary practice session. One learner, assuming the role of teacher and referencing a list, orally provided each Japanese translation, while their partner, with their sheet turned over, responded by providing the corresponding English expressions. These roles were then reversed to allow each learner to experience both positions. The participant in the teacher role was instructed to promptly provide the correct English expression if their partner was unable to respond, and to offer corrections when errors occurred. This vocabulary practice session typically lasted 2–3 minutes. In the post-task group, the same vocabulary practice activity was conducted after the completion of the communication tasks.

[3]All materials and instructions are available in the OSF link: https://doi.org/10.17605/OSF.IO/FWXTE

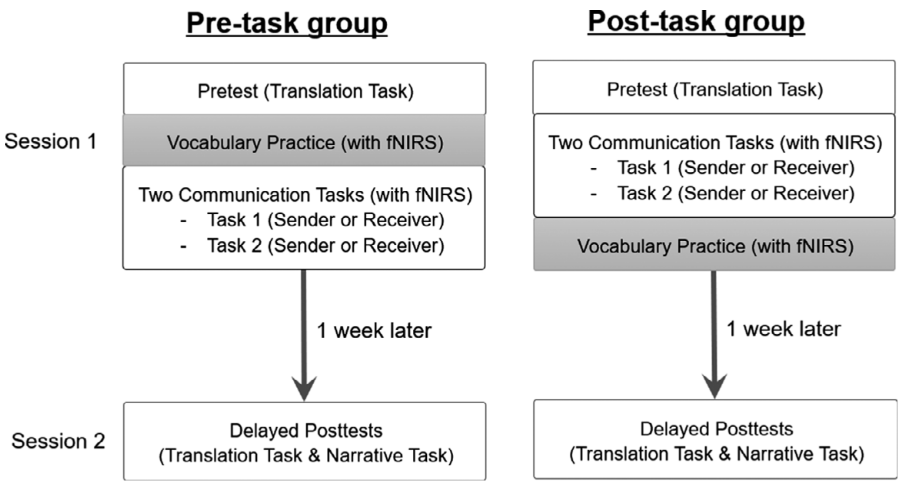

Figure 4. Experimental procedure.

During the communication tasks, participants were encouraged to use English exclusively and were allotted 20 minutes to complete each task. Importantly, while participants in the pre-task group had been told at the beginning of vocabulary practice with the instructions sheet that they would be learning "English phrases that will help with the communication task," they were not explicitly instructed when beginning the communication task to use these specific expressions. Both conditions received identical task instructions to ensure comparability. To control for potential order effects, we implemented a counterbalancing design across two factors: condition (pre-task vs. post-task) and task (Company Spy and Museum Thief). Across the two tasks, participants alternated roles, with each participant serving as sender in one task and receiver in the other.

One week after the first session, each learner returned to complete surprise delayed posttests (translation and narrative tasks) and an elicited imitation task to measure their L2 proficiency. The narrative task was administered only at the delayed posttest to avoid influencing the communication task performance in the first session. Similarly, immediate posttests were not administered to prevent potential positive practice effects on subsequent performance (Suzuki, 2017).

Data analysis

All analyses were conducted using R (R Core Team, 2022) with mixed-effects modeling. This approach allowed us to account for both fixed and random effects, thereby controlling for individual differences among participants and items (Baayen et al., 2008).

As accuracy scores were binary (1 or 0), we employed logistic mixed-effects models using the "lme4" package (Bates et al., 2015). For speech rate scores, we used linear mixed-effects models (LMMs) within the same package, as LMMs are well-suited for analyzing continuous data, such as speech rate.

To address RQ1 (the effect of vocabulary practice timing) and RQ2 (the predictive power of vocabulary practice performance), we employed a principled model selection procedure (Meteyard & Davies, 2020) to construct four models: two for the translation task (accuracy and speed measures) and two for the narrative task (accuracy and speed

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

measures). This procedure allowed us to systematically examine the effects of our key variables while controlling for potential confounders. We began by constructing a base model featuring Condition (pre-task vs. post-task) and several relevant covariates as fixed effects, with participants and items as random intercepts. The post-task condition was coded as 0 (reference level) because this group did not receive the vocabulary practice treatment prior to the communication tasks, with the pre-task condition coded as 1. For effect size, we report the odds ratio (OR) for the Condition effect, along with its 95% confidence interval (CI). The OR represents the change in odds of a higher accuracy score for the pre-task condition compared to the post-task condition, whereby OR > 1 indicates higher odds of accuracy in the pre-task condition, while OR < 1 indicates lower odds. Relevant R code is presented in Appendix F in Online Supplementary File.

The scores from the vocabulary practice phase (Vocabulary Practice), representing the total correct recalls of target expressions by the participants, were entered as a fixed effect. Additional covariates included pretest score on the translation test (Pretest), elicited imitation score (EIT), and partner contact frequency (Partner). These covariates were centered and included to control for individual differences in prior knowledge, language proficiency, and partner contact frequency that might influence lexical learning. Using the Variance Inflation Factor (VIF), multicollinearity assumption was checked for all models. As the VIF values for all predictors in our models were well below 10, multicollinearity was not a concern in our analyses.[4] We also tested an interaction between Condition and Vocabulary Practice performance, retaining it only if it significantly improved model fit based on likelihood ratio tests and AIC.

To address RQ3, we investigated the influence of pre-task vocabulary practice on IBS during subsequent communication tasks. For this purpose, brain activity in both members of each dyad was measured simultaneously using fNIRS (for the detailed neural data preprocessing procedure, see Appendix G in Online Supplementary File). Similar to the strategy described above, we conducted linear mixed-effects model analyses to examine differences between the post-task and the pre-task group. One mixed-effects model was constructed for the IBS during the two communication tasks (Task 1 and Task 2) as a within-subject fixed effect, while vocabulary practice, pretest, EIT, and partner served as covariates (for the R code, see Appendix H in Online Supplementary File). Based on the VIF scores, multicollinearity was not a concern in these analyses.

For RQ4, we extended these models to include six IBS predictors identified through the analysis addressing RQ3. Specifically, individuals' mean IBS scores for the mPFC and left LPFC for each learning block (i.e., the IBS for the mPFC for Vocabulary Practice, Task 1, and Task 2, as well as the IBS for the left LPFC for Vocabulary Practice, Task 1, and Task 2) were entered into the original mixed-effects models for RQ1 and RQ2. These IBS scores were the same as those used in the statistical model for RQ3. Using likelihood ratio tests, we compared these extended models to the original models to determine if the inclusion of IBS predictors significantly improved the model fit. The VIF scores confirmed that multicollinearity was not a concern in these analyses. The R code for the full models is presented in Appendix I in Online Supplementary File.

This comprehensive approach allowed us to rigorously test our hypotheses while accounting for the complex structure of our data and the potential interactions between

[4]Pretest and vocabulary practice scores were moderately correlated, but they are distinct as vocabulary practice score is the result of brief yet deliberate practice and is higher than the pretest score.

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

Table 2. Descriptive statistics for key variables by condition

	Pre-task Condition			Post-task Condition		
	M	95% CI	SD	M	95% CI	SD
[Covariates]						
Translation_Acc (Pretest)	29.66%	[26.35, 32.97]	10.20%	30.74%	[26.12, 35.36]	14.45%
Vocabulary Practice	55.55%	[50.75, 60.48]	16.82%	58.77%	[53.64, 63.66]	16.27%
[Posttest Measures]						
Translation_Acc (Posttest)	66.32%	[61.00, 71.64]	16.41%	63.86%	[58.28, 69.45]	17.47%
Translation_SpR (Posttest)	1.12	[1.07, 1.17]	0.14	1.10	[1.05, 1.15]	0.15
Narrative_Acc (Posttest)	60.16%	[53.32, 67.00]	21.38%	54.27%	[46.66, 61.88]	23.79%
Narrative_SpR (Posttest)	1.7	[1.53, 1.86]	0.52	1.71	[1.48, 1.94]	0.71

Note: CI = confidence interval; Acc = accuracy; SpR = speech rate.

variables. By using appropriate models for different types of outcome variables and employing a systematic model selection procedure, we ensured that our results were robust and interpretable within the context of our research questions.

Results

Descriptive statistics (Table 2) revealed similar baseline performance and vocabulary practice scores between conditions. Independent samples t-tests confirmed no significant differences in the pretest score on the translation task (Welch's $t(78) = 0.51$, $p = .61$, $d = 0.11$) or Vocabulary Practice performance ($t(78) = 0.88$, $p = .38$, $d = 0.20$). These results establish comparable starting points for both groups, allowing for meaningful comparisons of the timing manipulation's effects on subsequent task performance.

The effects of timing and vocabulary practice performance on lexical learning (RQ1 and RQ2)

Tables 3 through 6 show the full results of the logistic mixed-effects model for accuracy measures and of the linear mixed model for speech rates. In the remainder of this section, our interpretation will focus on the effects of Condition (RQ1) and Vocabulary Practice performance (RQ2).

Translation task

For the accuracy measure of the translation task (Table 3), the logistic mixed-effects model indicated a significant effect of Condition ($b = 0.343$, $p = .044$). The OR for Condition (pre-task vs. post-task) was 1.41 (95% CI [1.01, 1.97]), indicating that participants in the pre-task condition had 41% higher odds of achieving greater

Table 3. Logistic mixed-effects model results (accuracy in the translation task)

Predictor	*b* [95% CI]	SE	*z*	*p*
(Intercept)	0.96 [0.14, 1.79]	0.42	2.29	.02
Condition	0.34 [0.01, 0.68]	0.17	2.01	.04
Vocabulary Practice	0.08 [0.05, 0.11]	0.01	5.6	<.001
Pretest	0.19 [0.11, 0.26]	0.04	4.78	<.001
EIT	0.16 [−0.23, 0.56]	0.2	0.81	.42
Partner	−0.0002 [−0.11, 0.11]	0.06	−0	.99

Note: Random effects: Participant (SD = 0.40), Item (SD = 1.89).

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

Table 4. Linear mixed-effects model results (speech rate in the translation task)

Predictor	*b* [95% CI]	SE	*t*	*p*
(Intercept)	1.07 [0.95, 1.20]	0.06	17.51	<.001
Condition	0.03 [–0.02, 0.08]	0.03	1.15	.25
Vocabulary Practice	0.01 [0.01, 0.02]	0	3.93	<.001
Pretest	0.02 [0.004, 0.03]	0.01	2.70	.01
EIT	0.02 [–0.04, 0.08]	0.03	0.75	.46
Partner	–0.0003 [–0.02, 0.02]	0.01	–0.04	.97
Condition × Vocabulary Practice	–0.01 [–0.02, –0.003]	0	–2.74	.01

Note: Random effects: Participant (SD = 0.09), Item (SD = 0.27).

accuracy compared to those in the post-task condition. Vocabulary practice performance was a significant positive predictor of accuracy ($b = 0.081$, $p < .001$), with each unit increase in vocabulary practice performance associated with an 8% increase in the odds of higher accuracy (OR = 1.08, 95% CI [1.05, 1.12]).

For the speed measure in the translation task, the linear mixed-effects model (Table 4) revealed no significant effect of Condition ($b = 0.03$, 95% CI [–0.02, 0.08], $p = .25$). However, there was a significant positive effect of Vocabulary Practice performance on speech rate ($b = 0.01$, 95% CI [0.01, 0.02], $p < .001$) and a significant interaction between these variables ($b = –0.01$, 95% CI [–0.02, -0.003], $p = .01$). As shown in Figure 5, the interaction pattern suggests that higher vocabulary practice performance was associated with faster speech rates in the post-task condition only.

Narrative task

For the accuracy measure of the narrative task (Table 5), logistic mixed-effects model indicated a significant effect of Condition ($b = 0.428$, $p = .039$). The OR for Condition

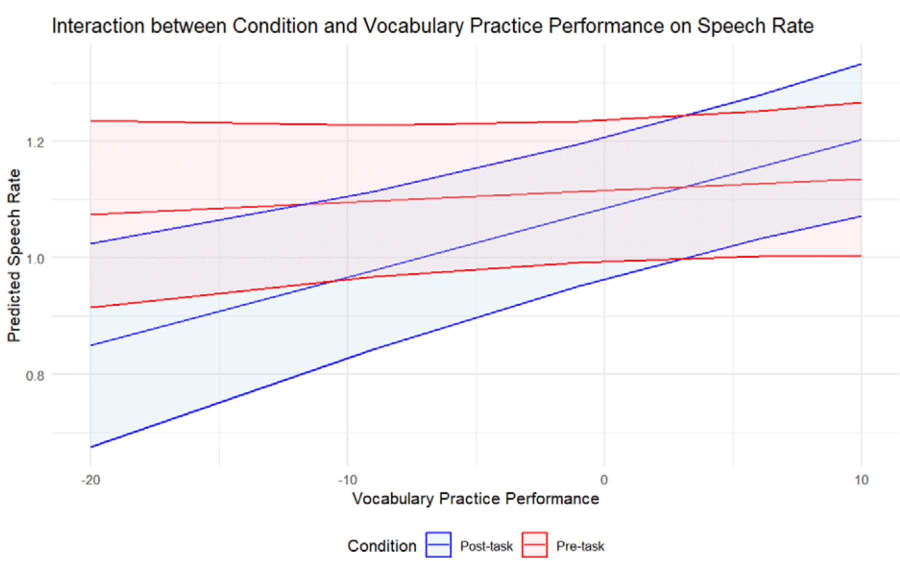

Figure 5. Effects of interaction between condition and vocabulary practice performance on speech rate in the translation task.
Note: The shaded areas represent 95% CIs.

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

Table 5. Logistic mixed-effects model results (accuracy in the narrative task)

Predictor	*b* [95% CI]	SE	*z*	*p*
(Intercept)	0.12 [–0.39, 0.63]	0.26	0.46	.65
Condition	0.43 [0.02, 0.83]	0.21	2.06	.04
Vocabulary Practice	0.10 [0.06, 0.13]	0.02	5.59	<.001
Pretest	–0.03 [–0.12, 0.06]	0.05	–0.71	.48
EIT	0.30 [–0.17, 0.77]	0.24	1.25	.21
Partner	–0.01 [–0.14, 0.12]	0.07	–0.19	.85

Note: Random effects: Participant (SD = 0.51), Item (SD = 0.74).

(pre-task vs. post-task) was 1.53 (95% CI [1.02, 2.30]), indicating that participants in the pre-task condition had 53% higher odds of achieving greater accuracy compared to those in the post-task condition. Vocabulary Practice performance remained a significant positive predictor of accuracy ($b = 0.099$, $p < .001$), with each unit increase in vocabulary practice performance associated with a 10% increase in the odds of higher accuracy (OR = 1.10, 95% CI [1.07, 1.14]). Other predictors did not show significant effects on accuracy in the narrative task.

The linear mixed-effects model for speech rate in the narrative task yielded results that were similar to those obtained in the translation task (Table 6). As in the translation task, the main effect of Condition was not significant ($b = 0.11$, 95% CI [–0.15, 0.37], $p = .41$), Vocabulary Practice performance showed a significant positive effect on speech rate ($b = 0.04$, 95% CI [0.01, 0.07], $p = .01$) and interacted with Condition ($b = –0.05$, 95% CI [–0.08, –0.01], $p = .01$). As illustrated in Figure 6, the interaction pattern for the speed measure in the narrative task was similar to that observed in the translation task.

The effects of timing on IBS (RQ3)

The mixed-effects model analysis revealed a significant main effect of Condition on the IBS in the left LPFC during communication tasks ($b = –0.02$, 95% CI [–0.04, –0.01], $p = .01$). Specifically, the post-task group exhibited higher mean IBS during communication blocks than the pre-task group. In contrast, no significant condition-related differences were observed in the mPFC (for full results, see Appendix J in Online Supplementary File).

Relationship between IBS and lexical learning (RQ4)

To address RQ4, we compared the extended models (including IBS predictors) to the original models (RQ1 and RQ2) using likelihood ratio tests. Results showed significant

Table 6. Linear mixed-effects model results (speech rate in the narrative task)

Predictor	*b* [95% CI]	SE	*t*	*p*
Intercept	1.60 [1.34, 1.87]	0.13	11.88	<.001
Condition	0.11 [–0.15, 0.37]	0.13	0.84	.41
Vocabulary Practice	0.04 [0.01, 0.07]	0.02	2.89	.01
Pretest	0.06 [–0.00, 0.11]	0.03	1.92	.06
EIT	0.22 [–0.08, 0.51]	0.15	1.45	.15
Partner	0.04 [–0.05, 0.12]	0.04	0.85	.40
Condition × Vocabulary Practice	–0.05 [–0.08, –0.01]	0.02	–2.64	.01

Note: Random effects: Participant (SD = 0.42), Item (SD = 0.32).

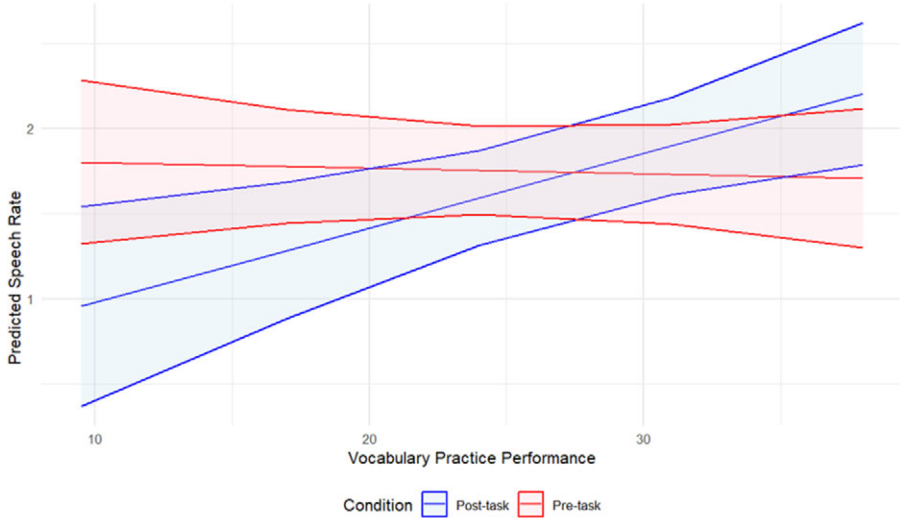

Figure 6. Effects of interaction between condition and vocabulary practice performance on speech rate in the narrative task.
Note: The shaded areas represent 95% CIs.

improvements in model fit for translation task accuracy ($\chi^2(6) = 20.60$, $p = .01$) and narrative task speech rate ($\chi^2(6) = 16.41$, $p = .01$), but not for translation task speech rate ($\chi^2(6) = 11.73$, $p = .07$) and narrative task accuracy ($\chi^2(5) = 6.25$, $p = .28$). We thus report the results for the two significant models and focus on the significant IBS predictors only (see full results in Appendix K in Online Supplementary File).

Translation task (accuracy measure)

Three IBS predictors exhibited significant effects on accuracy on the translation task: (a) the IBS in the mPFC during Vocabulary Practice ($b = 2.99$, 95% CI [0.08, 5.90], $p = .04$), (b) the IBS in the left LPFC in Task 1 ($b = 8.61$, 95% CI [3.68, 13.54], $p < .001$), and (c) the IBS in the mPFC in Task 2 ($b = 3.99$, 95% CI [0.57, 7.41], $p = .02$). These results suggest that increased IBS in specific regions and during particular tasks is associated with higher accuracy on the translation task.

Narrative task (speech rate)

The IBS in the mPFC during Vocabulary Practice showed a significant positive effect on speech rate in the narrative task ($b = 3.14$, 95% CI [0.75, 5.54], $p = .01$), suggesting that increased IBS in this area during vocabulary practice is associated with faster speech rate during the narrative task.

Discussion

The objective of the present study was to elucidate the effects of vocabulary practice timing on lexical learning and neural synchronization. Results pertaining to RQ1 support the hypothesis that pre-task vocabulary practice leads to greater learning than

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

post-task practice (H1). In the translation and the narrative task, participants who engaged in pre-task vocabulary practice demonstrated significantly higher accuracy compared to those in the post-task condition. This finding aligns with the view that initial development of declarative knowledge through form−meaning mapping, followed by subsequent contextual use of L2 items, enhances the consolidation of learned knowledge (Schmitt, 2008). From a TBLT perspective, the pre-task advantage may support Willis and Willis's (2007) argument that preparing learners with relevant linguistic resources can enhance task performance and subsequent acquisition.

The effect of vocabulary practice timing on speech rate was more nuanced. While there was no significant main effect of condition on speech rate in either task, we observed significant interactions between condition and vocabulary practice performance. This finding suggests that the impact of practice timing on production speed may be moderated by individual differences in vocabulary practice performance, the role of which is discussed next.

Contrary to our second hypothesis related to RQ2, the relationship between vocabulary practice performance and learning outcomes was not consistently stronger in the pre-task condition. Instead, we found that higher vocabulary practice scores predicted improved accuracy and speed in the translation as well as the narrative task across both conditions. This result underscores the importance of successful vocabulary practice, regardless of its timing, in facilitating lexical learning through interactive tasks.

The interaction between vocabulary practice performance and condition for speech rate measures reveals an intriguing pattern. In both tasks, higher vocabulary scores were associated with faster speech rates in the post-task condition, while this relationship was weaker or absent in the pre-task condition. One reviewer suggested this pattern might reflect a recency or priming effect, where vocabulary practice facilitated subsequent lexical access and production speed. This interpretation may warrant further investigations, given that the posttest took place one week after the vocabulary practice session. This significant lag between vocabulary practice and posttest sessions suggests that other factors are more likely at play.

Focusing on the impact of consolidation over time, for the post-task group, vocabulary practice was their final, focused exposure to the target items before the one-week delay. A strong performance in this session therefore likely predicted better consolidation and, consequently, faster retrieval a week later. In contrast, the pre-task group's initial practice performance did not predict their long-term retention. This may be because the subsequent communication task provided a crucial opportunity to automatize their knowledge through contextualized use. This interactive practice allowed them to enhance automaticity, making their final performance largely independent of how well they did in the initial practice session.

Regarding RQ3's question on how pre-task practice influences IBS, our exploratory H3 acknowledged different possibilities, given limited prior fNIRS research in interactive L2 learning. While we predicted increased left LPFC and decreased mPFC IBS, an alternative view considering reduced overall inter-personal coordination pointed to a potential decrease in both regions. The results revealed a complex picture: IBS in the left LPFC was significantly lower in the pre-task group compared to the post-task group, while no significant difference emerged in the mPFC. This finding, particularly the lower left LPFC IBS, runs counter to our primary prediction based on heightened attention to linguistic form. However, it is potentially consistent with the alternative possibility involving reduced interpersonal coordination, perhaps suggesting that activating specific lexical knowledge beforehand hinders, rather than facilitates, the real-time neural coordination underlying communicative language processing between partners. This observation could lend support to TBLT concerns that an excessive pre-task focus on linguistic forms can disrupt communication processes (e.g., R. Ellis et al., 2020; Skehan, 1998), possibly because learners' attention was divided. While prior work showed drawbacks of pre-task *grammar* instruction (R. Ellis et al., 2019), our study offers potential neural evidence for a similar effect with pre-task *lexical* instruction. The lack of significant effect in the mPFC did not clearly align with our prior predictions. Given the exploratory nature of H3 and these nuanced findings, the interpretations presented here remain speculative and warrant further investigation.

Finally, the results pertaining to RQ4 partially support the hypothesis that IBS (particularly in the left LPFC) predicts lexical learning outcomes by revealing complex relationships that vary across tasks and measures. In the translation task, increased IBS in the mPFC during vocabulary practice, in the left LPFC during the first task, and in the mPFC during the second task positively predicted accuracy. In the narrative task, the positive relationship between IBS in the mPFC during vocabulary practice and subsequent speech rate suggests that early synchronization in social-cognitive regions may set the stage for more fluent production in complex tasks, corroborating the findings yielded by previous studies showing that higher IBS enhances learning outcomes (e.g., Tan et al., 2023; Zhang et al., 2024).

Interestingly, the differential effects of IBS during Task 1 (left LPFC) and Task 2 (mPFC) on the translation accuracy may reflect the evolving nature of collaborative learning processes across sequential tasks. The left LPFC (language processing) appears more critical during initial collaborative interactions when participants are establishing communication patterns and negotiating meaning. In contrast, the mPFC (sociocognitive processing) becomes more important in subsequent collaborative tasks as participants develop shared understanding and more efficient communication strategies.

Taken together, these findings suggest that vocabulary acquisition through interactive tasks involves a dynamic interplay between language-specific neural networks and social-cognitive processes, with their relative contributions shifting as learners progress through sequential collaborative tasks.

Theoretical and practical implications

The interdisciplinary research findings reported in this work have both theoretical and practical implications. We revealed complex interactions among practice timing, IBS, and lexical learning, underscoring the need for more nuanced approaches to study optimal lexical support timing. The insights gained from IBS measures can offer a new perspective on the neural mechanisms underlying collaborative L2 learning from interactive tasks. We found that withholding vocabulary practice until communicative task completion is useful for promoting IBS. As engaging in interactive task prior to form-focused instruction could allow learners to grapple with communicative challenges, it may result in more intensive neural coupling and collaborative problem-solving, which also predicts lexical learning to some extent. Methodologically, we demonstrated that fNIRS can be a useful neuroimaging technique in elucidating the cognitive and social processes involved in L2 acquisition through task-based learning. Future research integrating behavioral and neural measures may provide more

https://doi.org/10.1017/S0272263125101290 Published online by Cambridge University Press

comprehensive models of how inter-brain dynamics support language learning in interactive contexts.

Practically, our findings highlight the need for teachers to carefully consider the timing of vocabulary practice based on their specific pedagogical goals and the nature of vocabulary knowledge they aim to develop in their students. Our analyses reveal that the timing of vocabulary practice can qualitatively alter the nature of the learning experience. According to Nation's (2007) four-strand principle, while communication tasks are typically categorized as meaning-focused output activities, our results suggest that incorporating vocabulary practice at different timings can shift learners' internal psychological orientation toward such activities. Pre-task vocabulary practice, while beneficial for enhancing vocabulary knowledge, may transform what is intended to be a meaning-focused output task into a more language-focused learning experience, as evidenced by reduced IBS during communication. As the timing of form-focused components critically influences how learners engage with the task, it would be helpful to consider not just which learning activities of four strands to combine, but also when to introduce them to maintain the intended pedagogical focus in class, aiming to strike a balance between lexical learning and communication processes.

Limitations and future directions

Several directions for future research can be proposed based on the present study. First, while the results indicate that pre-task vocabulary practice led to greater lexical learning gains, the current study did not analyze the interactional data produced during task completion itself. Examining the data—including the frequency and nature of target vocabulary usage (e.g., specific verb–noun combinations) as well as how dyads collaboratively approached the task—was beyond the scope of this paper. Such an analysis would be informative, because the target verb–noun combinations were selected to be task-*useful* rather than task-*essential.* This means that the lexical items could facilitate communication but their use was not strictly required, which aligns with the TBLT principle (R. Ellis et al., 2020). Analyzing the dyadic interactions is a valuable next step, potentially revealing relationships between specific interactional patterns, lexical use during the task, and lexical acquisition. Second, because the neural measures adopted in this study revealed different IBS patterns in the pre-task group, investigating learners' attention during communication tasks may help elucidate the effects of pre-task lexical instruction. Third, this cross-sectional laboratory experiment is considered an exploratory attempt to unveil the potential benefits and costs of pre-task vocabulary practice for task-based learning. As fNIRS technologies are increasingly being applied to investigate learning processes in classrooms (Tan et al., 2023, for a review), our laboratory study should be extended to classroom-based research with a longitudinal design, allowing the effects of vocabulary practice timing on lexical learning and communication processes to be monitored over time.

Conclusion

This study investigated the effects of vocabulary practice timing on lexical learning and IBS during interactive communication tasks, offering insights into the complex interplay among practice timing, neural processes, and lexical learning. The obtained results indicate that pre-task vocabulary practice is more conducive to lexical learning, as evidenced by improved accuracy in both translation and narrative tasks. Intriguingly, post-task vocabulary practice resulted in higher IBS in the left LPFC during communication tasks, suggesting that engaging in communication before formal vocabulary practice may foster greater neural synchronization through collaborative problem-solving. Furthermore, IBS in the brain regions responsible for social cognition and language processing during vocabulary practice and communication tasks predicted aspects of lexical learning, highlighting the importance of mutual collaboration in lexical learning through interaction.

Underscoring the importance of considering both cognitive and social aspects of language acquisition, the findings reported here have significant implications for TBLT and instructed SLA research. In particular, given that pre-task vocabulary practice enhances lexical learning but may potentially interfere with communication processes, a balanced approach in TBLT is required, carefully weighing the benefits of lexical preparation against the value of genuine communication.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263125101290.

Acknowledgements. We are grateful for the feedback we received on this work presented at the Vocab@-Maryland conference. This work is dedicated to the conference organizers from the Graduate Program in Second Language Acquisition at the University of Maryland, College Park—the first author's alma mater. Finally, our gratitude goes to Soma Sekiguchi and Atsushi Miura for their dedicated efforts in preparing this new experiment and collecting the data for this study. The study is funded by Grant–in–Aid for Scientific Research (KAKENHI) from Japan Society for the Promotion of Science (21H00550).

Data availability statement. The experiment in this article earned Open Data and Open Materials badges for transparent practices. The materials are available at https://osf.io/FWXTE/.

References

Altamimi, A., & Conklin, K. (2024). Effects of pre-reading study and reading exposure on the learning and processing of collocations. *TESOL Quarterly*, *58*(4), 1324–1346. https://doi.org/10.1002/tesq.3268

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

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

Cheng, J., Matthews, J., Lange, K., & McLean, S. (2023). Aural single-word and aural phrasal verb knowledge and their relationships to L2 listening comprehension. *TESOL Quarterly*, *57*(1), 213–241. https://doi.org/10.1002/tesq.3137

de la Fuente, M. J. (2006). Classroom L2 vocabulary acquisition: Investigating the role of pedagogical tasks and form-focused instruction. *Language Teaching Research*, *10*(3), 263–295. https://doi.org/10.1191/1362168806lr196oa

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge University Press.

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). Routledge.

Elgort, I., Wetering, R. v. d., Arrow, T., & Beyersmann, E. (2024). Previewing novel words before reading affects their processing during reading: An eye-movement study with first and second language readers. *Language Learning*, *74*(1), 78–110. https://doi.org/10.1111/lang.12579

Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 75–93). Routledge.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Ellis, R., & He, X. (1999). The roles of modified input and output in the incidental acquisition of word meanings. *Studies in Second Language Acquisition*, *21*(2), 285–301. https://doi.org/10.1017/S0272263199002077

Ellis, R., Skehan, P., Li, S., Shintani, N., & Lambert, C. (2020). *Task-based language teaching: Theory and practice*. Cambridge University Press.

Fotos, S., & Ellis, R. (1991). Communicating about grammar: A task-based approach. *TESOL Quarterly*, *25*(4), 605–628.

Gilbert, S. J., Spengler, S., Simons, J. S., Steele, J. D., Lawrie, S. M., Frith, C. D., & Burgess, P. W. (2006). Functional specialization within rostral prefrontal cortex (area 10): A meta-analysis. *Journal of Cognitive Neuroscience*, *18*(6), 932–948. https://doi.org/10.1162/jocn.2006.18.6.932

Godfroid, A. (2019). Sensitive measures of vocabulary knowledge and processing: Expanding nation's framework. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 433–453). Routledge.

González-fernández, B., & Schmitt, N. (2020). Word knowledge: Exploring the relationships and order of acquisition of vocabulary knowledge components. *Applied linguistics*, *41*(4), 481–505. https://doi.org/10.1093/applin/amy057

Gurzynski-Weiss, L. (2024). Domain-specific research methods in instructed second language acquisition: A next step for research integrity and impact. *Annual Review of Applied Linguistics*, *44*, 19–44. https://doi.org/10.1017/S0267190524000102

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Kelly, C., & Suzuki, Y. (2025). *The snoop detective school: Interactive tasks for english learners*. ABAX ELT Publishing.

Kelsen, B. A., Sumich, A., Kasabov, N., Liang, S. H. Y., & Wang, G. Y. (2022). What has social neuroscience learned from hyperscanning studies of spoken communication? A systematic review. *Neuroscience & Biobehavioral Reviews*, *132*, 1249–1262. https://doi.org/10.1016/j.neubiorev.2020.09.008

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied linguistics*, *29*(4), 694–716.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Li, P., & Jeong, H. (2020). The social brain of language: Grounding second language learning in social interaction. *npj Science of Learning*, *5*(1), 8.

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092. https://doi.org/10.1016/j.jml.2020.104092

Mora, J. C., & Levkina, M. (2017). Task-based pronunciation teaching and research: Key issues and future directions. *Studies in Second Language Acquisition*, *39*(2), 381–399. https://doi.org/10.1017/S0272263117000183

Nation, I. S. P. (2007). The four strands. *Innovation in Language Learning and Teaching*, *1*(1), 2–13. https://doi.org/10.2167/illt039.0

Nation, I. S. P. (2017). *The bnc/coca level 6 word family lists* (version 1.0.0) *[dataset]* (http://www.victoria.ac.nz/lals/staff/paul-nation.aspx

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press.

Newton, J. (2013). Incidental vocabulary learning in classroom communication tasks. *Language Teaching Research*, *17*(2), 164–187. https://doi.org/10.1177/1362168812460814

Nozawa, T., Sakaki, K., Ikeda, S., Jeong, H., Yamazaki, S., Kawata, K. H. d. S., Kawata, N. Y. d. S., Sasaki, Y., Kulason, K., Hirano, K., Miyake, Y., & Kawashima, R. (2019). Prior physical synchrony enhances rapport and inter-brain synchronization during subsequent educational communication. *Scientific Reports*, *9*(1), 12747. https://doi.org/10.1038/s41598-019-49257-z

Pellicer-Sánchez, A., Conklin, K., & Vilkaitė-Lozdienė, L. (2021). The effect of pre-reading instruction on vocabulary learning: An investigation of L1 and L2 readers' eye movements. *Language Learning*, *71*(1), 162–203. https://doi.org/10.1111/lang.12430

Quaresima, V., & Ferrari, M. (2019). Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: A concise review. *Organizational Research Methods*, *22*(1), 46–68. https://doi.org/10.1177/1094428116658959

R Core Team. (2022). R: A language and environment for statistical computing. Vienna, austria: R foundation for statistical computing. http://www.r-project.org/

Read, J. (2000). *Assessing vocabulary*. Cambridge University Press.

Saito, K., Uchihara, T., Takizawa, K., & Suzukida, Y. (2025). Individual differences in L2 listening proficiency revisited: Roles of form, meaning, and use aspects of phonological vocabulary knowledge. *Studies in Second Language Acquisition*, *47*(1), 26–52. https://doi.org/10.1017/S027226312300044X

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Springer.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921

Segalowitz, N. S. (2010). *Cognitive bases of second language fluency*. Taylor & Francis.

Siyanova-Chanturia, A., & Pellicer-Sanchez, A. (2018). *Understanding formulaic language: A second language acquisition perspective*. Routledge.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.

Suzuki, Y. (2017). The optimal distribution of practice for the acquisition of L2 morphology: A conceptual replication and extension. *Language Learning*, *67*(3), 512–545. https://doi.org/10.1111/lang.12236

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

Suzuki, Y. (2023). *Practice and automatization in second language research: Perspectives from skill acquisition theory and cognitive psychology*. Routledge.

Suzuki, Y., Maie, R., & Hui, B. (2025). Research timeline: Automatization in second language learning. *Language Teaching*, 1–20. Early View. https://doi.org/10.1017/S026144482500059X

Takizawa, K. (2024). What contributes to fluent L2 speech? Examining cognitive and utterance fluency link with underlying L2 collocational processing speed and accuracy. *Applied Psycholinguistics*, *45*(3), 516–541. https://doi.org/10.1017/S014271642400016X.

Tan, S. H. J., Wong, J. N., & Teo, W.-P. (2023). Is neuroimaging ready for the classroom? A systematic review of hyperscanning studies in learning. *Neuroimage*, *281*, 120367. https://doi.org/10.1016/j.neuroimage.2023.120367

Tyler, A. E., & Ortega, L. (2018). Usage-inspired L2 instruction: Some reflections and a heuristic. In A. E. Tyler, L. Ortega, M. Uno, & H. Park (Eds.), *Usage-inspired L2 instruction: Researched pedagogy* (Vol. *49*, pp. 315–321). John Benjamins Publishing Company.

Uchihara, T. (2022). Is it possible to measure word-level comprehensibility and accentedness as independent constructs of pronunciation knowledge? *Research Methods in Applied Linguistics*, *1*(2), 100011. https://doi.org/10.1016/j.rmal.2022.100011

Uchihara, T. (2025). Vocabulary and L2 listening. In M. Reed & J. Levis (Eds.), *The handbook of second language listening* (pp. 415–428). Wiley-Blackwell.

Uchihara, T., Saito, K., Kurokawa, S., Takizawa, K., & Suzukida, Y. (2025). Declarative and automatized phonological vocabulary knowledge: Recognition, recall, lexicosemantic judgement, and listening-focused employability of L2 words. *Language Learning*, *75*(2), 458–492. https://doi.org/10.1111/lang.12668

Webb, S. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *The Canadian Modern Language Review*, *65*(3), 441–470. https://doi.org/10.3138/cmlr.65.3.441

Willis, J., & Willis, D. (2007). *Doing task-based teaching-Oxford handbooks for language teachers*. Oxford University Press.

Wood, D. (2019). Classifying and identifying formulaic language. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 30–45). Routledge.

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge Uniersity Press.

Wu, S.-L., Tio, Y. P., & Ortega, L. (2022). Elicited imitation as a measure of L2 proficiency: New insights from a comparison of two L2 English parallel forms. *Studies in Second Language Acquisition*, *44*(1), 271–300. https://doi.org/10.1017/S0272263121000103

Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading?: A meta-regression analysis. *Studies in Second Language Acquisition*, *42*(2), 411–438. https://doi.org/10.1017/S0272263119000688

Zhang, P., & Graham, S. (2020). Learning vocabulary through listening: The role of vocabulary knowledge and listening proficiency. *Language Learning*, *70*(4), 1017–1053. https://doi.org/10.1111/lang.12411

Zhang, Y., Hu, Y., Ma, F., Cui, H., Cheng, X., & Pan, Y. (2024). Interpersonal educational neuroscience: A scoping review of the literature. *Educational Research Review*, *42*, 100593. https://doi.org/10.1016/j.edurev.2024.100593

Cite this article: Suzuki, Y., Nozawa, T., Uchihara, T., Nakamura, S., Miyazaki, A., & Jeong, H. (2025). Timing matters for interactive task-based learning: Effects of vocabulary practice on learning multiword expressions and neural synchronization. *Studies in Second Language Acquisition,* 47: 1018–1043. https://doi.org/10.1017/S0272263125101290