Studies in Second Language Acquisition (2026), 1–18
doi:10.1017/S0272263126102101

RESEARCH REPORT

Making task engagement visible and interpretable: A task-driven composite approach

Sachiko Nakamura[1] and Yuichi Suzuki[2]

[1]Tamagawa University, Japan and [2]Faculty of International Research and Education, Waseda University, Japan
Corresponding author: Sachiko Nakamura; Email: sachiko8266@gmail.com

(Received 06 January 2026 UTC; Revised 23 July 2026 UTC; Accepted 17 September 2026 UTC)

Abstract

This study investigates the effects of vocabulary support on engagement and lexical learning in paired interactive information-gap tasks. Sixty-six Japanese learners of English as a foreign language (EFL) performed the tasks either with or without access to task-relevant collocations. Engagement was operationalized through a task-driven composite approach, in which discourse-analytic indicators from transcribed interactions were summarized via principal component analysis. This resulted in three engagement patterns reflecting (a) use and negotiation of target collocations, (b) co-construction of task-relevant information and joint meaning making, and (c) task-maintenance behaviors. Bayesian regression and mediation analyses showed meaningful group differences in the engagement patterns but provided little evidence that these patterns mediated gains in productive knowledge of the target collocations, suggesting a complex relationship between task engagement and vocabulary knowledge development. Methodologically, the study illustrates how this approach can retain the visibility of engagement indicators, allowing greater interpretability of findings in relation to task design and instructional decision-making.

Keywords: information-gap task; lexical learning; task-driven composite approach; task engagement; TBLT

Introduction

Task-based language teaching (TBLT) is widely valued for fostering active learner engagement. Engagement, broadly understood as active participation and involvement in learning activities (Hiver et al., 2021; Philp & Duchesne, 2016), has been theorized as a prerequisite for successful second language (L2) acquisition (Dörnyei & Kormos, 2000) and gained increasing prominence in language learning research (Hiver et al., 2021), particularly within TBLT (for a comprehensive overview, see Hiver et al., 2025; Qiu, 2025).

While considerable empirical effort has been devoted to examining how tasks can promote engagement, less is understood about how task engagement relates to L2 learning outcomes (Qiu, 2025). In investigating such links, greater transparency in the

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

operationalization of engagement may help address methodological challenges highlighted in recent studies (Aoyama et al., 2024; Qiu, 2025).

Extending Suzuki and Nakamura (2026), the present study examines the role of engagement in mediating the relationship between vocabulary support and lexical learning in interactive information-gap tasks. In doing so, we propose a task-driven composite approach to operationalizing engagement, which supports pedagogically interpretable findings while retaining the visibility of the specific behaviors that constitute engagement.

Literature review

Engagement–learning link in task-based language teaching

To date, a substantial body of TBLT research has examined factors shaping task engagement (Qiu, 2025), including task features and task types (e.g., Lambert et al., 2016; Nakamura et al., 2020) as well as learner variables (e.g., Aubrey et al., 2020; Dao & McDonough, 2018). An emerging line of research has also begun to examine the outcomes of task engagement. Recent studies have investigated how engagement relates to L2 learning outcomes in relation to task-related factors, such as task type (Newton, 2013), content (Lambert & Aubrey, 2025), and design (Phan & Dao, 2023), as well as learner-related factors, such as self-reference (Qahl & Lambert, 2025) and task motivation (Hiver & Dao, 2025).

The present study contributes to these lines of inquiry by examining an under-explored feature of the TBLT framework, that is, explicit lexical support provided before and during task performance. It investigates how such support may shape learners' engagement in information-gap tasks and how engagement may relate to lexical learning. TBLT implementations vary in the extent to which linguistic support is provided before task performance. The inclusion of pre-task lexical support in the present study may align with task-supported language teaching (TSLT), which incorporates a brief focus on form prior to task performance (Li et al., 2016).

Pedagogically, TBLT frameworks are expected to offer teachers multiple options for task implementation. Understanding how support mechanisms shape learner engagement and subsequent learning can provide useful guidance for instructional design. To this end, findings from task engagement research should be pedagogically interpretable.

Operational challenges in task engagement research

To date, task engagement has commonly been researched based on Philp and Duchesne's (2016) multidimensional model, which conceptualizes engagement as a multifaceted construct involving behavioral, cognitive, social, and affective dimensions. Within this framework, task engagement is often operationalized through discourse-analytic indicators (i.e., indicators identified in learners' transcribed interaction during the task), which are assigned a priori to specific dimensions. For example, the number of words produced during the task is often used as an indicator of behavioral engagement, whereas instances of negotiation of meaning or form are typically interpreted as indicators of cognitive engagement (see Lambert & Aubrey, 2023, for a comprehensive review). Selected indicators are then pre-grouped according to the dimensions they are assumed to represent. Findings are then frequently summarized and discussed at the dimensional level.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

While the dimensional approach has generated valuable insights and supported the accumulation of findings across studies through meta-analyses and systematic reviews (Hiver et al., 2021; Okunuki & Kashimura, 2024; Qiu, 2025), two aspects of this methodology warrant further attention. The first concerns the visibility and consistency of task engagement operationalizations. As demonstrated in recent reviews (Qiu, 2025), operationalizations of engagement dimensions vary considerably across studies, with different sets of indicators used to represent individual dimensions and some inconsistencies noted (Aoyama et al., 2024). As a result, summarizing findings at the dimensional level may obscure important variation in the specific indicators used to operationalize engagement, making it difficult to build a cumulative and concrete understanding of task engagement across studies.

The second aspect concerns pedagogical interpretability when findings are reported primarily at the dimensional level. For instructors working within TBLT frameworks, understanding how learners engage in concrete task behaviors (e.g., actively using target lexical items through interaction) may be more informative than knowing whether a task increased or decreased a particular engagement dimension (e.g., behavioral engagement). Similarly, descriptions of specific engagement behaviors that are closely linked to task features may offer more direct insight into instructional design decisions.

These considerations suggest the potential value of an approach that maintains the clarity of the indicators used to operationalize engagement while summarizing them in a contextually interpretable and pedagogically meaningful manner.

The current study

This study aims to contribute to task engagement research in two ways. First, it aims to enhance the methodological visibility and pedagogical interpretability of engagement operationalization by adopting a task-driven composite approach in which task-specific discourse-analytic indicators are summarized through principal component analysis (PCA).

Second, it seeks to shed light on the under-researched link between explicit vocabulary support, engagement in interactive information-gap tasks, and task-based L2 lexical learning outcomes. To this end, the following research questions (RQs) are addressed:

RQ1. How do students engage in interactive information-gap tasks, as reflected in observed engagement indicators?

RQ2. To what extent does task engagement mediate the effect of vocabulary support on task-based lexical learning?

Methods

The dataset used in the present study is drawn from Suzuki and Nakamura (2026), who examined the effects of explicit vocabulary support during interactive information-gap tasks on L2 lexical learning and affective responses to tasks (i.e., task emotions, perceived task difficulty, and perceived partner collaborativeness). The present study uses the lexical learning component of this dataset and extends the original design by incorporating a new analysis of learners' engagement behaviors based on discourse data from the task interactions.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

Participants

The participants were 66 Japanese university students[1] (38 male, 28 female; mean age = 19.42 years, *SD* = 1.19) enrolled in six intact English communication courses. The cohort represented a range of academic majors, primarily liberal arts (25.0%), agriculture (25.1%), and international business (20.6%). Their English proficiency, based on TOEIC[2] test scores converted to the Common European Framework of Reference for Languages (CEFR; Educational Testing Service, 2025), was predominantly at the A2 level (77.3%), with smaller cohorts at the A1 (6.1%) and B1 (10.6%) levels; TOEIC scores were unavailable for four students.

The six classes were assigned to one of two conditions: a task-only group (three classes, n = 36) or a vocabulary-support group (three classes, n = 30). To ensure proficiency was balanced within the interactive dyads, students were paired based on their scores from the junior English Minimal Test (jMET), a validated dictation test of general English proficiency (Goto et al., 2010).

An independent-samples Welch t-test, using jMET scores, showed no significant proficiency difference between the vocabulary-support and control groups ($t(60.28)$ = 0.51, p = .61). A Shapiro–Wilk test indicated that jMET scores were normally distributed (W = .99, p = .80).

Materials

Information-gap tasks. Two interactive information-gap tasks, titled "Company Spy" and "Museum Thief," were adapted from the TBLT textbook by Kelly and Suzuki (2024). As shown in Figure 1, learners were organized into pairs, with one student acting as the "sender" and the other as the "receiver." The sender, provided with a series of pictures depicting a spy's or thief's actions with corresponding time stamps, described the events to the receiver, who then had to plot the information onto a corresponding map of the location. The objective was for the pair to collaboratively reconstruct the sequence of events. To ensure each learner experienced both roles, students acted as the sender for one task and the receiver for the other. Learners were instructed to interact in English during the tasks, and 20 min were allocated to complete each task.

Target collocation list. In the vocabulary-support condition, learners were provided with a list of the useful collocations corresponding to key actions in the task materials (Appendix A in the Online Supplementary Material). The list served as optional linguistic support to facilitate task completion, and learners were free to use their own lexical resources in addition to the provided expressions. The list consisted of 16 verb–noun collocations, with eight items designated for each task (Company Spy and Museum Thief). These items (e.g., *climb a ladder*, *release sleeping gas*) were selected for their functional relevance to successful task completion rather than on strict statistical grounds, aligning with a view of formulaic language that prioritizes contextual utility (Wood, 2019). The constituent verbs and nouns were high-frequency words drawn from the top 1,000 word families of the British National Corpus (BNC)/Corpus of Contemporary American English (COCA) word family lists (Nation, 2017), and the verb–noun combinations were reviewed by a native English speaker, who confirmed

[1]Two students from Suzuki and Nakamura (2026) were excluded from this study due to missing data.

[2]The Test of English for International Communication (TOEIC) is a widely used standardized English proficiency test that assesses listening and reading skills in workplace-oriented and everyday communication contexts, with scores ranging from 10 to 990.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

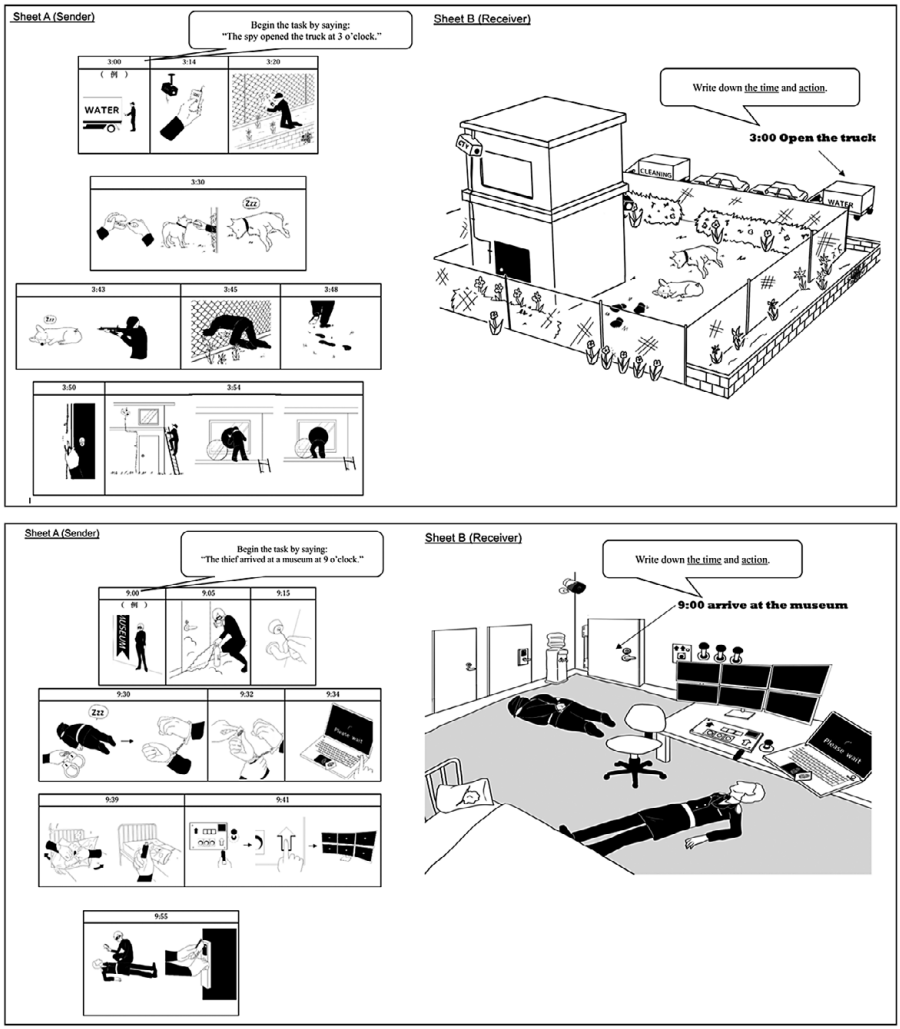


Figure 1. Information-gap tasks: Company Spy (top) and Museum Thief (bottom).

them as natural and appropriate expressions for describing the target scenes in the task materials. A previous study that employed longer versions of the same information-gap tasks found that Japanese university learners did not consistently produce these specific combinations (Suzuki et al., 2025). Taken together, we anticipated that although the learners in the current study might recognize the individual words within the verb–noun combinations, they would struggle to produce these specific combinations spontaneously, and that vocabulary support would therefore lead to greater use of the target collocations during task performance. Although some combinations may be less common in A2-level materials, they were considered suitable pedagogical targets in this study.

Lexical knowledge test. To measure gains in collocational knowledge, a written production test was administered before and after the tasks. The pretest and posttest were identical in content, each containing the 16 target collocations and 8 distractor

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

items, with the item order randomized between administrations. For each item, learners were given a Japanese phrase and required to write the corresponding English collocation in blank spaces, cued with the first letter and word length (e.g., カメラのスイッチを切る = (s _ _ _ _ _) (o _ _) a (c _ _ _ _ _)). This form-recall format was designed to assess productive knowledge of the target forms. The complete test is available in Appendix B in the Online Supplementary Material.

Operationalization of task engagement: A task-driven composite approach

In this study, we operationalized task engagement through a task-driven composite approach. This approach begins with identifying discourse-analytic indicators that reflect learners' active participation and involvement in the task. The focus on behaviors draws on theoretical views that emphasize action as a core characteristic of engagement (Hiver et al., 2021; Oga-Baldwin, 2019). Next, the identified indicators were subjected to PCA to group co-occurring indicators into composite behavioral patterns. These composite behavioral patterns, commonly referred to as *components* in PCA practice, are alternative forms of the original variables (DeVellis & Thorpe, 2021), representing the unique engagement manifested in the specific task context. As such, the term, composite, in this study denotes the summarization of multiple observable task-specific discourse-analytic indicators into broader engagement patterns (for detailed discussions of composite modeling, see, e.g., Alamer et al., 2024).

Methodologically, this operationalization offers contextualized measures of task engagement, supporting the task-dependent nature of engagement emphasized in the literature (Aoyama et al., 2024; Hiver et al., 2025). The grouping of indicators was treated as exploratory, as the study aimed to identify patterns of co-occurring engagement behaviors in this particular dataset. To preserve the visibility of the behaviors that constitute the operationalization of engagement, the components are described primarily in terms of the specific indicators.

Analysis

To investigate how engagement was represented through observable interactional behaviors during the task (RQ1), we employed a two-phase qualitative-quantitative procedure based on the task-driven composite operationalization of task engagement. The first phase was qualitative, aimed at identifying discourse-analytic engagement indicators that were characterized by the specific task demands while being informed by the literature on task engagement. As such, the analysis proceeded in the following steps. First, the learners' audio-recorded task performances were transcribed using a verbatim transcription method.[3] Next, three sets of transcripts (i.e., a total of 12 task performances by six students) were read through and qualitatively examined to identify engagement indicators based on the literature on task engagement (Aoyama et al., 2024; Philp & Duchesne, 2016; Qiu, 2025; Reinders & Nakamura, 2021) and previous studies adopting discourse-analytic measures of task engagement (e.g., Lambert & Aubrey, 2023; Lambert et al., 2016; Nakamura et al., 2020; Trofimovich et al., 2021). For the

[3]The transcripts indicated that interaction was predominantly conducted in English, with occasional L1 use. Upon closer examination, we decided not to exclude episodes containing L1 as these instances were closely embedded within interactional sequences that were otherwise conducted in English, making it impractical to separate L1 and L2 segments without disrupting the continuity of the interaction.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

indicators identified in the transcripts, we further examined their roles in relation to key elements of this study, such as the interactive nature of information-gap tasks and typical discourse sequences, as well as the provision of collocation lists. Through multiple discussions and reexaminations, eight indicators were established as initial measures of task engagement. Accordingly, a coding manual was developed, detailing the definitions of these indicators and outlining the coding rules and instructions.

The second phase was primarily quantitative, coding the occurrences of these indicators in the transcripts and using PCA to group indicators parsimoniously. First, using the manual and coded transcripts, we trained an independent coder, a senior undergraduate student majoring in applied linguistics.[4] Then, the coder and the first author independently coded transcripts from seven pairs (> 20% of the dataset) to assess the inter-rater reliability. Krippendorff's alpha was .99, indicating near-perfect agreement (Krippendorff, 2004). For collocation use, which involved not only identifying occurrences but also assigning partial scores (e.g., 0.5 when only the verb or noun was used), the intraclass correlation coefficient [ICC(2,1)] was computed. The result revealed an ICC of 1.00 (95% CI [1.00, 1.00]), indicating perfect agreement (Koo & Li, 2016). The remaining data were coded by the independent coder. During the coding process, any unclear cases were discussed and resolved collaboratively, ensuring consistent application of the coding scheme. The occurrences of indicators for each pair were normalized by the time on task to compute engagement scores for individual students.

The engagement scores were then entered into a PCA. To determine the number of components to retain, we used a parallel analysis based on principal components, which provides a more robust criterion than the traditional eigenvalue-greater-than-one rule (DeVellis & Thorpe, 2021; Patil et al., 2008). An oblique rotation, Oblimin, was applied to allow for correlations among components. Simultaneously, the correlation matrix was examined to understand relationships among indicators (see Appendix C in the Online Supplementary Material). To examine the stability of the final PCA loading pattern, we conducted a bootstrap-based stability check using 1,000 resamples. Tucker's congruence coefficient was used to assess the similarity of the component loading patterns, following the guidelines proposed by Lorenzo-Seva and Berge (2006): Values from .85 to .94 indicate fair similarity, whereas values higher than .95 indicate that components can be considered equivalent.

To create composite scores for each engagement component, we applied a regression-based weighting approach using the component loadings from the PCA. Specifically, each indicator was multiplied by its corresponding loading (i.e., weight) on the relevant principal component, and the weighted scores were summed to generate composite scores. The resulting composite scores were standardized (*z*-scores) for subsequent analysis.

With regard to RQ2, when exploring how engagement mediated the effect of vocabulary support on lexical learning, we conducted Bayesian[5] regression analyses

[4]The independent coder underwent extensive training over 2 months, including seven training and calibration meetings with the first author, totaling approximately 15 hr, as well as independent study and practice coding. In these meetings, the coder and the first author reviewed indicator definitions, compared independently coded transcripts, discussed discrepancies, and refined the coding rules to ensure consistent application of the manual.

[5]The present study employed Bayesian modeling, as it allows flexible estimation of complex models and yields full posterior distributions for all parameters, supporting transparent interpretation of parameter uncertainty (McElreath, 2020; Yuan & MacKinnon, 2009).

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

followed by mediation models using the *brms* package (Bürkner, 2017) in R (version 4.4.2).

In the first regression model, the three engagement components derived from PCA served as outcome variables, with vocabulary support as the predictor and L2 proficiency as a covariate. The second model examined the direct path from vocabulary support to lexical learning, and the third model examined direct paths from each engagement component to lexical learning. Pre-task vocabulary scores and grand-mean-centered jMET scores (i.e., L2 proficiency) were included as covariates in these models.

The final analyses consisted of Bayesian mediation models, with vocabulary support as the predictor, an engagement component as the mediator, and post-task vocabulary scores as the outcome. Pre-task vocabulary scores and grand-mean-centered jMET scores were included as covariates in all paths. All variables were standardized before analysis.

Bayesian regression and mediation models were estimated using the *brms* package (Bürkner, 2017) with the *cmdstanr* backend. Models were fitted using four chains with 4,000 iterations each (1,000 warmup).

Default weakly informative priors implemented in *brms* were used. Flat priors were specified for regression coefficients, while weakly informative Student-*t* priors were used for intercepts and residual standard deviations, constraining parameters to plausible ranges while remaining robust to outliers (Gelman et al., 2008). Full prior specifications for all models are provided in Appendix D in the Online Supplementary Material.

Convergence diagnostics indicated a satisfactory model fit ($\hat{R} = 1.00$; high bulk and tail effective sample sizes), with no divergent transitions. Posterior predictive checks (Appendix E in the Online Supplementary Material) indicated that the models reproduced the observed data distributions reasonably well.

Indirect effects were estimated using the mediation() function from the *bayestestR* package (Makowski et al., 2019). Following Gignac and Szodorai (2016), standardized regression coefficients (β) of 0.10, 0.20, and 0.30 were interpreted as small, medium, and large effects, respectively. We applied these benchmarks to relationships between continuous variables. For the paths that involved the condition being a binary predictor (0 = task only, 1 = vocabulary support), effects were interpreted based on the direction and credibility of the posterior estimates.

Results

Engagement in interactive information-gap tasks (RQ1)

Examination of the task performance transcripts identified eight discourse-analytic indicators reflecting how students engaged in the interactive information-gap tasks. Figure 2 presents the indicators with descriptions and representative excerpts.

A prominent feature of the interaction was the presence of negotiation moves, as learners worked to resolve communication difficulties and complete the task by identifying missing information and collaboratively filling the information gaps. Such negotiation moves were oriented toward addressing communication breakdowns (i.e., negotiation of meaning; Pica et al., 1989), linguistic form (i.e., negotiation of form; Lyster & Ranta, 1997), and task-related information (i.e., negotiation of content; Rulon & McCreary, 1986).

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

Indicator	Description	Representative excerpt
Negotiation of meaning, content, and form for clarification	Seeking clarification of a partner's prior utterance due to uncertainty, lack of clarity, or linguistic or communicative difficulty.	A：Um, flower near alumi, alumi wall. B：Alumi wall? Oh. A：cutting. B：Alumi wall cutting? A：Yeah.
Negotiation of meaning, content, and form for elaboration	Asking for elaboration to obtain additional information needed to complete the task.	B：Um, door. A：Yeah. B：Door, umm..... A：1 or 2? B：One. Ah, white one, white door. A：Yeah.
Negotiation of meaning, content, and form, involving target collocations	Seeking clarification or elaboration, where the negotiation involves the use of target collocations.	A：Umm, sleeping, a thief is, umm, put a hand cuff, hand cuffs for. B：Hand pats? A：Hand cuffs. B：Ah, Hm? Hm? Hand? A：Hand cuff. B：Ah.
Target collocation use	Using target collocations, categorized as full (both verb and noun used) or partial (only the verb or the noun used).	A：3:54. Climbed, climb up a ladder.
Peer completion	Completing a partner's utterance when they pause, use Japanese, or produce incomplete speech; this may also involve adding to or refining the partner's expression.	B：Black, OK. Ummm... time, 3..... A：Ah, 40. B：3:40. OK.
Response confirmation	Confirming that information has been heard or understood, typically demonstrated by the information receiver.	A：Umm, no neckless dog, shot a gun. B：Ah, OK. OK.
Echoing	Repeating information uttered by the information provider, typically demonstrated by the information receiver.	A：2. Umm, collect a fingerprint. B：Collect a fingerprint. Hmmm, OK.
Checking in	Checking whether the partner is following or has understood so far, typically demonstrated by the information provider.	A： A card key. Hm? OK. B：OK? OK? A：OK.

Figure 2. A list of engagement indicators with descriptions and representative excerpts.
Note: A coding manual is provided in Appendix F in the Online Supplementary Material.

The negotiation moves were further observed to serve two distinct purposes: clarification and elaboration, reflecting the communicative demands of the information-gap tasks. Clarification moves were observed when learners sought to resolve uncertainty regarding a partner's prior utterance, whereas elaboration moves arose when additional detail was required to complete the task.

When a partner's utterance was incomplete or unclear, learners sometimes supplied or refined the utterance, a behavior identified as peer completion. It should be noted that similar behaviors have been described as turn completion (Nakamura et al., 2020) and responsiveness (Trofimovich et al., 2021) in previous task engagement research.

Given the direct relevance of the target collocations provided as part of the instructional support, learners' use of target collocations was treated as a distinct engagement indicator, rather than being subsumed under overall lexical production.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

Similarly, negotiation moves involving target collocations were distinguished from negotiation moves that did not involve target collocations.

Another indicator identified in the transcripts was response confirmation, in which learners signaled that their partner's utterance had been heard or understood, typically through brief acknowledgments such as "ok" or "yes." Echoing was also observed, whereby learners repeated information previously provided by their partner. These indicators reflected learners' responsiveness (Trofimovich et al., 2021) and functioned as small but critical moves that helped learners to proceed from one information gap to the next. Checking-in behaviors were additionally identified when learners verified that their task partner was following or had understood the information provided.

Descriptive statistics for the eight indicators are illustrated in Figure 3, with full values reported in Appendix G in the Online Supplementary Material. While the indicators varied in overall frequency, some tendencies were observable across conditions. For example, target collocation use and collocation-related negotiation moves appeared more frequently in the vocabulary-support condition, whereas peer completion and clarification-oriented negotiation moves appeared more frequently in the task-only condition.

When these scores were analyzed using PCA, the initial PCA revealed that one of the eight indicators, echoing, cross-loaded on two components (loadings = .41 and .69), making it difficult to assign it to a single coherent engagement component. After its removal, the subsequent PCA with the remaining seven indicators yielded a clear and interpretable solution comprising three rotated components (hereafter RC1, RC2, and RC3), as shown in Table 1. The three-component solution explained 73.09% of the total variance. Communalities ranged from .55 to .84, indicating that most indicators were well represented by the retained components. Inspection of the scree plot also supported this three-component solution (see Appendix H in the Online Supplementary Material). Furthermore, the component correlation matrix indicated minimal associations among the components (see Appendix I in the Online Supplementary Material), suggesting that the identified engagement patterns were distinct. The bootstrap-based stability check further supported the overall stability of the three-component PCA solution (see Appendix J in the Online Supplementary Material).

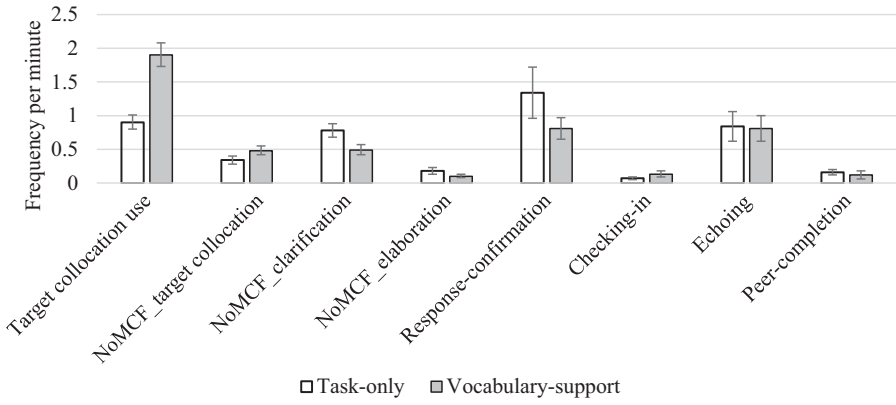


Figure 3. Descriptive statistics for task engagement indicators.
Note: Error bars indicate 95% confidence intervals. NoMCF = negotiation of meaning, content, and form.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

Table 1. Principal component loadings for engagement indicators

	RC1	RC2	RC3	Communality
Target NoMCF	.91			.84
Target collocation	.83			.79
Response confirmation	–.55		.51	.65
NoMCF clarification		.86		.75
NoMCF elaboration		.68		.60
Peer completion		.62		.55
Checking in			.90	.82

Note: RC = Rotated component. NoMCF = negotiation of meaning, content, and form. Only loadings ≥ .40 are displayed; the full loading matrix is reported in Appendix K in the Online Supplementary Material.

Taken together, the PCA results indicated that the indicators formed three distinguishable patterns of co-occurring engagement behaviors, as described below:

- **RC1** reflected collocation-focused engagement, characterized by the use of target collocations in interaction. This component consisted of two indicators: target collocation use and negotiation of meaning, content, and form involving target collocations.
- **RC2** reflected co-construction of task-relevant information and joint meaning making, characterized by interactional behaviors through which learners collaboratively clarified and elaborated task content. This component consisted of three indicators: negotiation of meaning, content, and form for clarification; negotiation of meaning, content, and form for elaboration; and peer completion.
- **RC3** reflected task-maintenance behaviors that supported the progression of the interaction. This component consisted of two indicators: response confirmation and checking in.

It should be noted that we retained response confirmation only in RC3, as the negative loading of response confirmation on RC1 suggested that this behavior moved in the opposite direction from the other two indicators in that component.

Mediating role of engagement in the effect of vocabulary support on lexical learning (RQ2)

RC1 (target collocation use + collocation-related negotiation). Initial Bayesian regression analyses showed that vocabulary support had a credible, positive effect on lexical learning (β = 0.77, 95% credible interval [CrI] [0.46, 1.08]). Vocabulary support also had a credible, positive effect on RC1 (β = 1.45, 95% CrI [1.11, 1.79]). In turn, RC1 strongly predicted lexical learning, with a large effect size (β = 0.30, 95% CrI [0.13, 0.47]).

In the final Bayesian mediation model (Figure 4), vocabulary support retained a credible positive effect on RC1 (β = 1.45, 95% CrI [1.10, 1.80]). Vocabulary support also retained a credible positive direct effect on lexical learning even after including RC1 in the model (β = 0.71, 95% CrI [0.25, 1.17]). The indirect effect of vocabulary support on lexical learning via RC1 was weak and non-credible (β = 0.06, 95% CrI [–0.27, 0.40]), suggesting little evidence that RC1 mediated the relationship between vocabulary support and lexical learning.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

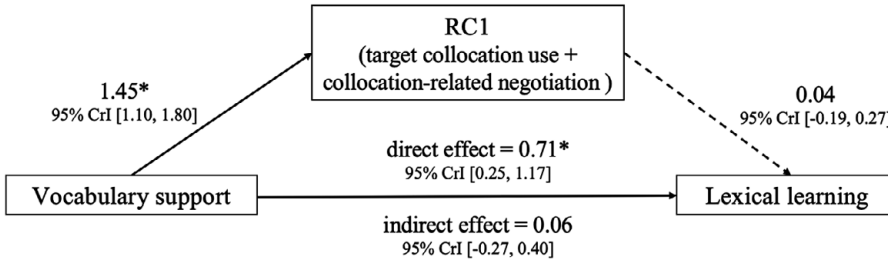


Figure 4. Mediation model for RC1.
Note: Standardized regression coefficients (β) are presented. Asterisks indicate credible effect (95% CrI does not include zero). Dotted lines represent non-credible pathways.

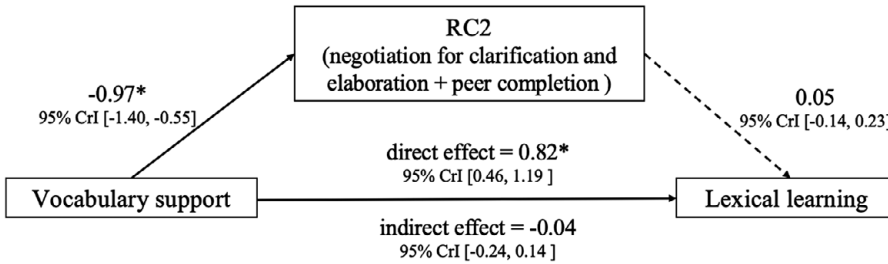


Figure 5. Mediation model for RC2.
Note: Standardized regression coefficients (β) are presented. Asterisks indicate credible effect (95% CrI does not include zero).

RC2 (negotiation for clarification and elaboration + peer completion). Bayesian regression analyses showed that vocabulary support had a credible, negative effect on RC2 (β = –0.97, 95% CrI [–1.40, –0.54]). While RC2 showed a small positive relationship with lexical learning, the relationship was not credible (β = 0.05, 95% CrI [–0.14, 0.23]).

In the final Bayesian mediation model (Figure 5), vocabulary support retained a credible negative effect on RC2 (β = –0.97, 95% CrI [–1.40, –0.55]) and a positive effect on lexical learning (β = 0.82, 95% CrI [0.46, 1.19]). The indirect effect was weak and non-credible (β = –0.04, 95% CrI [–0.24, 0.14]), showing little evidence for mediation.

RC3 (response confirmation + checking in). Vocabulary support showed no credible effect on RC3 (β = –0.45, 95% CrI [–0.93, 0.04]). Similarly, no credible pathway was shown from RC3 to lexical learning (β = –0.16, 95% CrI [–0.34, 0.02]). Accordingly, mediation was not tested.

Discussion

Engagement in interactive information-gap tasks (RQ1)

Our first RQ aimed to identify and describe the ways in which learners engaged in interactive information-gap tasks with greater transparency and interpretability. To this end, task engagement was operationalized using discourse-analytic indicators from the transcripts. Of the eight indicators initially indentified, seven were summarized into

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

three engagement patterns using PCA: RC1 (target collocation use + collocation-related negotiation), RC2 (negotiation for clarification and elaboration + peer completion), and RC3 (response confirmation + checking in).

Notably, these engagement patterns do not appear to map neatly onto the commonly established dimensions of engagement. RC1 seems to reflect features commonly associated with both the behavioral dimension (i.e., production of target lexical items) and the cognitive dimension (i.e., target collocation–related negotiation moves). Similarly, RC2 includes characteristics associated with the social dimension (i.e., peer completion) as well as the cognitive dimension (i.e., negotiation for clarification and elaboration).

On the other hand, the three engagement patterns appear to align closely with the key characteristics of the tasks used in this study. Learners used concrete vocabulary to precisely describe and exchange information contained in the task materials (RC1). Even when they did not possess knowledge of the precise lexical items, they negotiated and co-constructed meaning and information to resolve discrepancies in understanding (RC2). Learners also maintained interaction through response confirmation and checking-in moves that supported the progression of the task (RC3). Taken together, these findings further echo the notion that engagement is context-dependent and shaped by the specific demands of the task (Aoyama et al., 2024; Hiver et al., 2025; Philp & Duchesne, 2016).

Vocabulary support, engagement, and lexical learning (RQ2)

Our second RQ explored the role of engagement in mediating the relationship between vocabulary support and lexical learning. The findings showed that the provision of vocabulary support shaped how learners engaged in the tasks. Learners who received a list of task-relevant collocations engaged in the tasks by actively utilizing these lexical items in completing the tasks (RC1) while maintaining the progress of the interaction (RC3). Learners without the vocabulary support also demonstrated task-maintenance behaviors (RC3), but their engagement was more strongly characterized by joint meaning making and co-construction behaviors, such as negotiation for clarification and peer completion (RC2). This contrast is illustrated in Figure 6, which presents a comparative excerpt showing how a vocabulary-support pair directly used the target collocation, whereas a task-only pair relied on paraphrasing and negotiation to convey the intended meaning.

Despite these distinct engagement patterns observed between the two groups, none of these patterns credibly mediated the effect of vocabulary support on lexical learning.

Pair ID #5 [Task-Only]	# Pair ID #38 [Vocabulary-Support]
A：Next, hmm, I look spy's... spy's, spy's foot, foot..... B：Ah, this? Ah, three, three-foot... A：Yeah, yeah, yeah. B：OK, OK, shadow. Three-foot shadow? A：Yeah. B：OK, OK, OK. A：Three-foot shadow.	C：Leave footprint. D：Leave, leave to? C：Leave footprints. D：What time? C：3:48. Hmmm. D：Leave foot print. OK

Figure 6. Excerpts from task-only and vocabulary-support groups.
Note: Target collocation = leave footprints.

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

The result showed that learners' gains in productive knowledge of the target collocations, as measured by the form-recall test, were primarily attributable to the vocabulary support itself and that engagement did not appear to mediate this relationship.

While this result may seem counterintuitive, particularly for the vocabulary-support condition, where learners frequently used the target collocations in performing the tasks, this result importantly highlights the different facets and modalities of lexical learning (Nation, 2001). The vocabulary tests in this study primarily assessed written lexical knowledge, and the potential contribution of collocation-oriented engagement may not have been fully captured by this modality. Had auditory lexical knowledge also been assessed, the analysis might have revealed a relationship between collocation-oriented engagement and lexical development. This account also highlights the importance of input-assessment congruence in vocabulary research (Kaderoğlu, 2026). Additionally, this interpretation may help explain differences between the present findings and previous task engagement studies that have reported positive relationships between engagement and learning, whose operationalization of L2 learning was more closely aligned with the linguistic processes activated during task performance (e.g., Hiver & Dao, 2025; Suzuki et al., 2026).

Another possible account for these non-credible mediation results relates to the relatively short treatment period (40 min across the two information-gap tasks). Although learners in the vocabulary-support condition used the target collocations frequently during interaction (approximately twice per minute), the absolute learning gains were modest, averaging 2.92 collocations out of the 16 targets, compared with 0.50 in the task-only group (Suzuki & Nakamura, 2026). It therefore remains uncertain whether this amount of task interaction was sufficient for interactional engagement to translate into gains in productive knowledge as measured by the form-recall test. Moreover, the restricted range of learning gains may have limited the statistical detection of an indirect effect. These considerations call for replication studies with more extensive interventions (e.g., designs incorporating task repetition) that provide repeated opportunities for learners to encode and consolidate the target collocations.

From a pedagogical perspective, the contrasting engagement patterns across the two conditions are themselves informative. Although learners without vocabulary support did not demonstrate substantial gains in productive knowledge of the target collocations, they showed higher frequencies of interactional engagement behaviors, such as peer completion and negotiation of meaning, form, and content. Such behaviors may, in principle, encourage collaborative learning and contribute to the development of interactional competence.

Lastly, the findings illustrating the contextualized manifestation of engagement and its mediating role should be understood within the specific implementation adopted in the current study, in which explicit vocabulary support was provided prior to task performance—an approach best framed as TSLT. In stronger TBLT approaches, where linguistic support is provided more reactively, such as in response to learners' emergent communicative needs during or after task performance, the engagement patterns observed across groups may have been less distinct. This, in turn, could have influenced the mediating role of engagement.

Limitations

We acknowledge several limitations of the present study. First, the data were drawn from a relatively small sample at a single instructional site, reflecting the classroom-based nature of this study. Although smaller sample sizes may be considered adequate

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

for PCA when analyses indicate strong correlations among variables and a limited number of components (Tabachnick & Fidell, 2013), the results should be interpreted with appropriate caution. Similarly, while Bayesian analyses allow for more stable estimation with smaller samples (Yuan & MacKinnon, 2009), and while careful attention was paid to the specification of priors and interpretation of credible intervals, caution is warranted when generalizing the findings beyond the present context. The current exploratory findings should be replicated with larger samples across multiple instructional sites to establish their generalizability (Vitta et al., 2022).

Second, we administered only an immediate posttest. The absence of a delayed posttest prevented us from examining the long-term retention of the observed learning gains. In addition, as discussed above, the written form-recall test may not have fully captured the aspects of collocational knowledge developed through oral interaction (Nation, 2001). Future research should therefore incorporate delayed posttests as well as modality-congruent measures, such as auditory form recall (Kaderoğlu, 2026), to trace the development and retention of different aspects of collocational knowledge.

Third, while the present study focused on the role of vocabulary support in shaping engagement patterns, learners' engagement is likely to be co-determined by learner-internal variables (e.g., language aptitude, personality) and interactional and contextual variables (e.g., pair dynamics, L1 use, classroom climate). Future research may therefore benefit from examining how multiple learner- and task-related variables jointly shape engagement behaviors in interactive tasks, as well as incorporating multimodal indicators of engagement.

Finally, we retained instances of L1 use when coding engagement indicators as they were closely embedded in interactional discourse that was otherwise conducted in the L2. However, further investigation into the role of L1 use, such as through additional categories of engagement indicators (e.g., L2-only negotiation moves vs. negotiation moves involving code-switching), could provide richer insights into task engagement, particularly in instructed second language learning contexts where learners typically share the same L1 background.

Conclusion

This study examined how explicit vocabulary support shapes learners' engagement behaviors and task-based lexical learning in interactive information-gap tasks. By operationalizing engagement through discourse-analytic indicators, the study identified distinct patterns of engagement that reflect how learners collaboratively work toward task completion. The findings contribute to growing research exploring the role of engagement in language learning by demonstrating how external linguistic support may shape both interactional behaviors and learning outcomes within TBLT contexts.

Methodologically, this study illustrates the value of a task-driven composite approach to operationalizing engagement, which retains the visibility of the discourse-analytic indicators underlying engagement patterns. Because the approach identifies engagement patterns as they are manifested in a specific task and labels them in terms of concrete behaviors, findings can be communicated in a form that is both methodologically transparent and pedagogically accessible. For example, reporting that learners actively used and negotiated target collocations (RC1) conveys more actionable information to practitioners than reporting that learners demonstrated a higher level of behavioral engagement.

Pedagogically, the findings offer implications for task design and implementation in EFL contexts. When the instructional goal is for learners to use and acquire specific

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

target phrases, providing a pre-task collocation list appears effective: Learners with such support used the target collocations more frequently and made greater gains in productive knowledge. When the goal is to maximize negotiation-rich interaction, withholding pre-task lexical support may be preferable, as learners without the list engaged more in clarification-oriented negotiation and peer completion, drawing on their own linguistic resources. Given that no engagement pattern credibly mediated learning in this study, however, such design choices are best understood as shaping the interactional processes through which learning opportunities emerge, rather than as directly determining learning outcomes.

Taken together, these findings contribute to ongoing efforts to better understand how engagement functions within task-based interaction and how task implementation choices may support meaningful language use and learning.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263126102101.

Data availability statement. The dataset for this study is available on the Open Science Framework (OSF) at https://osf.io/zhc47.

Acknowledgments. We express our deepest gratitude to Yee-Ching Hon for her dedicated and invaluable contribution throughout the engagement coding process.

Funding. This work was supported by JSPS KAKENHI Grant Numbers 25K16369 and 25K00479.

Competing interests. The authors declare no competing interests.

References

Alamer, A., Schuberth, F., & Henseler, J. (2024). When and how to use confirmatory composite analysis (CCA) in second language research. *Studies in Second Language Acquisition*, *46*(2), 597–616. https://doi.org/10.1017/s0272263124000020

Aoyama, T., Yamazaki, J. S., Nakamura, S., Vuogan, A., An, H., Kim, C. J., & Al-Hoorie, A. H. (2024). Conceptualization and operationalization in L2 task engagement research: Taking stock and moving forward. *Language Teaching*, *57*(4), 1–5. https://doi.org/10.1017/s026144482400020x

Aubrey, S., Lambert, C., & Leeming, P. (2020). The impact of first as opposed to second language pre-task planning on the content of problem-solving task performance. *Language Teaching Research*, *26*(5), 867–892. https://doi.org/10.1177/1362168820917844

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Dao, P., & McDonough, K. (2018). Effect of proficiency on Vietnamese EFL learners' engagement in peer interaction. *International Journal of Educational Research*, *88*, 60–72. https://doi.org/10.1016/j.ijer.2018.01.008

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage.

Dörnyei, Z., & Kormos, J. (2000). The role of individual and social variables in oral task performance. *Language Teaching Research*, *4*(3), 275–300. https://doi.org/10.1177/136216880000400305

Educational Testing Service. (2025). *Mapping the TOEIC® tests on the Common European Framework of Reference for Languages* [PDF]. https://www.ets.org/pdfs/toeic/toeic-mapping-cefr-reference.pdf

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-aoas191

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Goto, K., Maki, H., & Kasai, C. (2010). The minimal English test: A new method to measure English as a second language proficiency. *Evaluation & Research in Education*, *23*(2), 91–104. https://doi.org/10.1080/09500791003734670

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2021). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, *28*(1), 201–230. https://doi.org/10.1177/13621688211001289

Hiver, P., & Dao, P. (2025). From task motivation to L2 learning: Understanding links through learners' task engagement. *TESOL Quarterly*, *59*(S2), S24–S59. https://doi.org/10.1002/tesq.3410

Hiver, P., Dao, P., & Yamazaki, J. S. (2025). L2 task engagement: A research agenda. *Language Teaching*, 1–23. https://doi.org/10.1017/S0261444825101043

Kaderoğlu, K. (2026). Aligning input and assessment: Pictorial measures in audiovisual vocabulary research. *Research Methods in Applied Linguistics*, *5*(1), Article 100288. https://doi.org/10.1016/j.rmal.2025.100288

Kelly, C., & Suzuki, Y. (2024). *The SNOOP detective school: Interactive tasks for English learners*. ABAX.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30*(3), 411–433. https://doi.org/10.1111/j.1468-2958.2004.tb00738.x

Lambert, C., & Aubrey, S. (2023). Discourse analytic methods. In C. Lambert, S. Aubrey, & G. Bui (Eds.), *The role of the learner in task-based language teaching* (pp. 113–124). Routledge.

Lambert, C., & Aubrey, S. (2025). Learner-generated content, task engagement, affective response, and memory on technology-mediated conversation tasks. *Studies in Second Language Learning and Teaching*, *15*(3), 479–500. https://doi.org/10.14746/ssllt.42079

Lambert, C., Philp, J., & Nakamura, S. (2016). Learner-generated content and engagement in second language task performance. *Language Teaching Research*, *21*(6), 665–680. https://doi.org/10.1177/1362168816683559

Li, S., Ellis, R., & Zhu, Y. (2016). Task-based versus task-supported language instruction: An experimental study. *Annual Review of Applied Linguistics*, *36*, 205–229. https://doi.org/10.1017/S0267190515000069

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: Negotiation of form in communicative classrooms. *Studies in Second Language Acquisition*, *19*(1), 37–66. https://doi.org/10.1017/S0272263197001034

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), Article 1541. https://doi.org/10.21105/joss.01541

McElreath, R. (2020). *Statistical rethinking: A Bayesian course with examples in R and Stan* (2nd ed.). CRC Press. https://doi.org/10.1201/9780429029608

Nakamura, S., Phung, L., & Reinders, H. (2020). The effect of learner choice on L2 task engagement. *Studies in Second Language Acquisition*, *43*(2), 428–441. https://doi.org/10.1017/s027226312000042x

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

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

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

Oga-Baldwin, W. L. Q. (2019). Acting, thinking, feeling, making, collaborating: The engagement process in foreign language learning. *System*, *86*, Article 102128. https://doi.org/10.1016/j.system.2019.102128

Okunuki, A., & Kashimura, Y. (2024). Student engagement and academic achievement in second language learning: A meta-analysis. *JACET Journal*, *68*, 71–90. https://doi.org/10.32234/jacetjournal.68.0_71

Patil, V. H., Singh, S. N., Mishra, S., & Donavan, D. T. (2008). Efficient theory development and factor retention criteria: A case for abandoning the "eigenvalue greater than one" criterion. *Journal of Business Research*, *61*(2), 162–170. https://doi.org/10.1016/j.jbusres.2007.05.008

Phan, H. L. T., & Dao, P. (2023). Engagement in collaborative writing: Exploring learners' control of task content and text quality. *International Journal of Applied Linguistics*, *33*(2), 242–259. https://doi.org/10.1111/ijal.12462

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, *36*, 50–72. https://doi.org/10.1017/s0267190515000094

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press

Pica, T., Holliday, L., Lewis, N., & Morgenthaler, L. (1989). Comprehensible output as an outcome of linguistic demands on the learner. *Studies in Second Language Acquisition*, *11*(1), 63–90. https://doi.org/10.1017/S027226310000783X

Qahl, T., & Lambert, C. (2025). Personal investment in language processing: The role of self-reference in initial lexical acquisition. *TESOL Quarterly*, *59*(S2), S60–S87. https://doi.org/10.1002/tesq.3402

Qiu, X. (2025). Communicative task engagement in second language learning and teaching: A scoping review. *Language Learning Journal*, *53*(6), 1–26. https://doi.org/10.1080/09571736.2025.2502733

Reinders, H., & Nakamura, S. (2021). Engagement. In S. Mercer & T. Gregersen (Eds.), *The Routledge handbook of psychology of language learning* (pp. 137–148). Routledge. https://doi.org/10.4324/9780429321498-14

Rulon, K. A., & McCreary, J. (1986). Negotiation of content: Teacher-fronted and small-group interaction. In R. R. Day (Ed.), *Talking to learn: Conversation in second language acquisition* (pp. 182–199). Newbury House.

Suzuki, Y., & Nakamura, S. (2026). Pre-task vocabulary support enhances lexical learning but dampens positive emotions: Interactive task implementation in English-as-a-foreign-language classroom. *TESOL Quarterly*, *60*(3), 1112–1142. https://doi.org/10.1002/tesq.70054

Suzuki, Y., Nozawa, T., Nakamura, S., Uchihara, T., Miyazaki, A., & Jeong, H. (2026). Same behavior, different brain: A triangulated neurocognitive re-evaluation of task engagement in TBLT. *Research Methods in Applied Linguistics*, *5*(3), Article 100331. https://doi.org/10.1016/j.rmal.2026.100331

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*(4), 1018–1043. https://doi.org/10.1017/S0272263125101290

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Trofimovich, P., Tekin, O., & McDonough, K. (2021). Task engagement and comprehensibility in interaction. *Journal of Second Language Pronunciation*, *7*(3), 435–461. https://doi.org/10.1075/jslp.21006.tro

Vitta, J. P., Nicklin, C., & McLean, S. (2022). Effect size–driven sample-size planning, randomization, and multisite use in L2 instructed vocabulary acquisition experimental samples. *Studies in Second Language Acquisition*, *44*(5), 1424–1448. https://doi.org/10.1017/S0272263121000541

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

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian mediation analysis. *Psychological Methods*, *14*(4), 301–322. https://doi.org/10.1037/a0016972

Cite this article: Nakamura, S., & Suzuki, Y. (2026). Making task engagement visible and interpretable: A task-driven composite approach. *Studies in Second Language Acquisition*, 1–18. https://doi.org/10.1017/S0272263126102101

https://doi.org/10.1017/S0272263126102101 Published online by Cambridge University Press