

Self-Regulated Learning and L2 Vocabulary Knowledge in Japanese University EFL Learners: Form Recognition, Meaning Recognition, and Proficiency-Level Differences



JESLA | Journal of the
European Second
Language Association

COLLECTION:
VOCABULARY
STUDIES AND THE
PSYCHOLOGY OF
LANGUAGE LEARNING:
TOWARDS GREATER
INTERDISCIPLINARY
COLLABORATION

RESEARCH

MEGUMI SUGITA

SIWON PARK

*Author affiliations can be found in the back matter of this article

WHITE ROSE
UNIVERSITY PRESS
Universities of Leeds, Sheffield & York

ABSTRACT

Vocabulary knowledge underpins second language (L2) development, but how much self-regulated learning contributes may depend on the aspect of lexical knowledge measured. We examined how vocabulary-specific self-regulated learning (SRLvoc) relates to two recognition-based aspects of vocabulary knowledge, form recognition and meaning recognition, in 267 first-year Japanese university English as a foreign language (EFL) students. Participants completed a 15-item SRLvoc scale adapted from Tseng et al. (2006), an 80-item Japan Association of College English Teachers (JACET) 8000-based vocabulary test (Levels 2–6) with separate form- and meaning-recognition sections, and the Test of English as a Foreign Language Institutional Testing Program (TOEFL ITP) as an index of receptive and grammatical proficiency. SRLvoc correlated more strongly with meaning recognition ($r = .41$) than with form recognition ($r = .18$); structural equation modeling indicated that this contrast rests primarily on zero-order correlations, because overlap between the two recognition factors made the path to meaning recognition unstable. Band-level regressions showed minimal SRLvoc-related variance at the highest frequency band and somewhat stronger associations in several mid-frequency meaning-recognition bands, and estimates were comparable across proficiency groups. SRLvoc is thus most evident when vocabulary knowledge is treated as recognition knowledge differentiated by retrieval direction rather than as a single composite score. We discuss these findings in relation to the integration of goal setting, monitoring, and spaced review within routine vocabulary instruction.

CORRESPONDING AUTHOR: Megumi Sugita

Kanda University of
International Studies, Japan
sugita-m@kanda.kuis.ac.jp

KEYWORDS:

L2 vocabulary; self-regulated learning (SRL); form recognition; meaning recognition; L2 proficiency; structural equation modeling (SEM)

TO CITE THIS ARTICLE:

Sugita, M., & Park, S. (2026). Self-Regulated Learning and L2 Vocabulary Knowledge in Japanese University EFL Learners: Form Recognition, Meaning Recognition, and Proficiency-Level Differences. *Journal of the European Second Language Association*, 10(1), 131–146. DOI: <https://doi.org/10.22599/jesla.164>

1. INTRODUCTION

Vocabulary size is one of the most reliable predictors of L2 performance (Park et al., 2026; Stæhr, 2008; Zhang & Zhang, 2022); however, learners receiving comparable instruction often develop markedly different lexicons. After several years of studying word lists, reading graded readers, and preparing for exams, some Japanese university students acquire a broad and stable vocabulary, while others plateau with smaller vocabularies. Proficiency, the amount of English exposure learners receive, and their opportunities to use the language explain much of this variation, but not all of it (Sun et al., 2023; Ueno & Takeuchi, 2025). A further factor is how learners manage their study time: how they plan it, maintain their focus, and prevent it from being eroded by boredom and distraction. These behaviors are the focus of research on self-regulated learning (SRL).

SRL refers to the processes through which learners set goals for their own study, monitor and adjust their cognition and motivation, and manage the conditions under which they work to ensure that learning continues over time (Oxford, 2016; Tseng et al., 2006). Applied to vocabulary acquisition, SRL encompasses decisions such as when to study new words, how often to review them, and what to do when motivation wanes. For this study, the key point is that SRL is not a single trait. Tseng et al. (2006) modeled self-regulating capacity in vocabulary learning (SRCvoc) as a set of distinguishable controls over commitment, metacognition, satiation, emotion, and the study environment. Subsequent construct validation research supports treating the instrument as a composite of several capacities rather than as a single undifferentiated scale (Alamer et al., 2024). We adopt this multidimensional perspective and refer to the construct, as operationalized for vocabulary learning in this study, as SRLvoc, representing it through three components: emotion and satiation control, goal regulation and commitment, and environmental and cognitive control. Collapsing these components into a single score, as is still common when SRL is entered as one predictor, can obscure the possibility that they relate differently to various aspects of vocabulary knowledge. The rationale for this three-component structure, which emerged from a pilot analysis, is given in Section 3.2.3. This conceptualization directly informs the statistical model introduced later: the same three components serve as indicators of a single common SRLvoc factor.

Importantly, how the relationship between SRL and vocabulary manifests in a dataset also depends on how vocabulary is measured. Recognition and recall impose different cognitive demands; frequency-banded tests sample various segments of the lexicon, and proficiency differences influence what learners can do with a given word (Akase, 2021; Laufer & Goldstein, 2004; McDonald & Asaba, 2015; Uchihara et al., 2022). Both vocabulary outcomes in the present study are multiple-choice recognition measures. Therefore, any association we report should be interpreted as a link to the steady exposure, review, and attention that maintain form-meaning connections retrievable under recognition support, rather than as evidence that SRLvoc drives productive use or deeper lexical control.

With this scope in mind, we examine how SRLvoc relates to the recognition-based vocabulary knowledge of Japanese first-year university students and whether this relationship varies depending on the direction of recognition (form versus meaning), the frequency level of words, and the learner's proficiency. We combine an adapted SRLvoc scale with a Japan Association of College English Teachers (JACET) 8000-based test that distinguishes form recognition from meaning recognition, and use Test of English as a Foreign Language Institutional Testing Program (TOEFL ITP) scores as an external measure of proficiency. The research questions are presented following the literature review.

2. LITERATURE REVIEW

2.1. CONCEPTUALIZING SRL IN L2 VOCABULARY

Self-regulation views learning as an active process managed by students rather than something that passively happens to them. Students decide what to focus on, regulate their effort and mood, and organize the conditions under which they study (Oxford, 2016; Tseng et al., 2006). In vocabulary acquisition, this manifests in small, everyday actions: setting aside time to learn new words, revisiting them before they are forgotten, persevering with a list even when it becomes tedious, and choosing how to review words. Three types of instruments have been used to

measure self-regulation: some are based on general models of strategic self-regulation, others on inventories of vocabulary learning strategies, and some on scales specifically designed for self-regulation in vocabulary learning (Alamer et al., 2024; Kudo et al., 2015; Tseng et al., 2006).

The instrument we use is an adapted version of Tseng et al.'s (2006) SRCvoc scale; throughout, we refer to the construct it measures as SRLvoc and to the instrument as the SRLvoc scale. Tseng et al. identified five facets of regulation: commitment, metacognition, satiation, emotion, and environmental control. We retain the multidimensional concept but consolidate the items into three components. This decision rests partly on the factor structure observed in a pilot analysis, reported in Section 3.2.3, and partly on the correspondence of the three components to the affective, goal-directed, and contextual aspects of regulation. Emotion and satiation control involves managing the stress and boredom commonly associated with vocabulary study. Goal regulation and commitment encompasses setting vocabulary goals, striving to achieve them, and avoiding procrastination; this component incorporates the goal-directed and metacognitive functions from Tseng et al.'s five original facets. Environmental and cognitive control pertains to organizing one's surroundings and maintaining concentration. These three components are particularly relevant to vocabulary learning because vocabulary acquisition relies on repeated, spaced encounters, which can be disrupted by frustration, satiation, procrastination, and distracting environments. The scale items specifically address these issues: recognizing boredom and responding to it, studying until a goal is achieved, and arranging the environment to facilitate rather than hinder learning. The full list of items, organized by component, is provided in Appendix A.

The question of how the scale should be modeled remains open. Using confirmatory composite analysis, Alamer et al. (2024) demonstrated that the SRCvoc functions as a composite constructed from several regulatory capacities rather than a single underlying trait. Interpreted this way, SRL in vocabulary learning is a configuration of emotion and satiation control, goal regulation and commitment, and environmental and cognitive control, rather than a one-dimensional disposition. A second perspective treats SRL as an upstream resource, in the sense that regulation sits earlier in the causal chain than any particular strategy and supplies the capacity that makes consistent strategy use possible. In mediated models, general self-regulation shapes more proximal motivational beliefs and strategy choices, which in turn influence vocabulary outcomes (Li & Wang, 2024). Both perspectives distinguish the capacity from specific strategies and allow for both direct and indirect effects on vocabulary. We adopt a similar approach, using a multidimensional measure and conceptualizing SRLvoc as a single common factor with the three components as its indicators. Our outcomes are recognition tests, so we keep the interpretation narrow: SRLvoc here can support the stability of form-meaning links sufficient for recognition, but it does not address productive command.

2.2. HOW STRONGLY DOES SRL PREDICT VOCABULARY KNOWLEDGE?

Current research suggests that the association between self-regulation and vocabulary knowledge is modest but meaningful. Ueno and Takeuchi's (2025) meta-analysis of correlational studies reports an average correlation of approximately .24 between vocabulary strategy use and vocabulary outcomes. Proficiency and exposure typically explain more variance. In Sun et al.'s (2023) multivariate analysis, both regulation and strategy use still accounted for unique variance when entered simultaneously. Additionally, Alamer et al. (2024) report stable, moderate correlations between the SRCvoc composite and vocabulary achievement across different institutions.

Intervention studies are less straightforward. Training learners in self-regulation or providing them with an app that supports it has resulted in greater vocabulary gains compared to standard instruction (Chen et al., 2024; Deng & Trainin, 2023; Fatemipour & Najafgholikhah, 2015). However, these gains tend to be smaller for weaker learners. We read this as a sign that self-regulation needs a base of vocabulary and motivation to work on, although the primary studies do not test this explanation directly. The studies also lack a common metric, employing different SRL scales, vocabulary assessments, and participant cohorts. Consequently, a single overall figure representing the SRL-vocabulary relationship is of limited value. Instead, the more pertinent questions concern where this relationship holds and for whom.

2.3. FORM RECOGNITION AND MEANING RECOGNITION AS VOCABULARY MEASURES

What a test measures depends on the task it requires the learner to perform. Vocabulary tasks are typically categorized along two dimensions simultaneously: recall versus recognition and form versus meaning. A form task requires access to the L2 word form, while a meaning task requires access to the meaning associated with it (Laufer & Goldstein, 2004; Meara, 1996; Nation, 2001; Schmitt, 2010). Combining these two dimensions results in four task types (Table 1): form recall (producing the L2 word), meaning recall (producing the L1 meaning), form recognition (selecting the L2 word given an L1 cue), and meaning recognition (selecting the L1 meaning given an L2 word). General findings from this framework indicate that recognition tasks are easier than recall tasks and that retrieving meaning is easier than retrieving form (Laufer & Goldstein, 2004; Sasao, 2008; Webb, 2008).

	RECALL	RECOGNITION
Form	Supply the L2 word	Select the L2 word
Meaning	Supply the L1 meaning	Select the L1 meaning

Table 1 Classification of vocabulary knowledge by recall/recognition and form/meaning retrieval.

We focus on two types of recognition tasks. Form recognition requires identifying the English form given a Japanese cue, while meaning recognition involves identifying the Japanese meaning given an English word. The practical rationale for using recognition rather than recall tasks is given in Section 3.2.2.

We also had reason to expect the two directions to behave differently. Recognizing meaning from an L2 form depends on how often a learner has encountered the word through reading and review. Self-regulation extends this record of encounters by sustaining review over time, for example through regular, spaced sessions (Nakata, 2015). In contrast, recognizing a form from an L1 cue depends more on how firmly the spelling has been encoded and on deliberate L1-to-L2 retrieval practice (Nakata, 2016). A measure of broad regulatory capacities captures these processes only indirectly, because it does not ask about task-specific retrieval practice. Therefore, we expected SRLvoc to align more closely with meaning recognition. This expectation is consistent with the strategy literature: input-oriented strategies, of which guessing from context and dictionary consultation are the best-documented examples, are associated more with recognition and receptive measures than with productive ones (Sevigny et al., 2024; Uchihara et al., 2022). Because recognizing a meaning from an L2 form presupposes some stability of form-level knowledge, the model also specifies a directional path from form recognition to meaning recognition (Section 3.4).

Both directions of recognition have previously been examined using structural equation modeling (Park et al., 2020). By modeling form recognition and meaning recognition as latent factors, each derived from performance across multiple frequency levels rather than a single observed score, we can separate the shared variance between the two abilities from the noise inherent in any single set of items. This approach also allows the SRLvoc comparison to be conducted while holding all other variables in the model constant. This is the approach we adopt.

2.4. VOCABULARY LEVEL AND FREQUENCY EFFECTS

Most vocabulary tests are organized by word frequency for a good reason: frequency roughly reflects the order in which words are learned (Nation, 2001), even though it predicts knowledge of individual words only moderately well (Hashimoto, 2021). Common words are encountered early and often, appearing first; rarer words are encountered less frequently and appear later. A frequency-based list transforms this concept into a measurement tool by selecting items from successive bands so that a learner's profile across these bands indicates the extent of their lexicon. In Japan, the standard list of this kind is the JACET 8000, compiled by the Japan Association of College English Teachers. It ranks the 8,000 most useful word families for Japanese learners and divides them into eight bands of 1,000 words each, with Level 1 being the most frequent and Level 8, the least. In this paper, we use band (or frequency band) for these divisions, retaining Level 2 through Level 6 as the JACET labels for specific bands, and band-level analysis means examining performance band by band rather than as a single total.

Learners' progress in knowledge across frequency bands is uneven. In a longitudinal study of Japanese secondary students, distinct clusters of strategy use corresponded with different growth trajectories at various proficiency levels (Akase, 2021), suggesting that self-regulation may facilitate vocabulary development in band-specific ways. Frequency also influences how learners cope with unknown words. In a bilingual vocabulary size test, guessing behavior varied according to both word frequency and learner proficiency, affecting the size estimate depending on whether scoring criteria were strict or lenient (McDonald & Asaba, 2015). How learners respond to this kind of uncertainty has also drawn attention in its own right, most recently as a motivational element in vocabulary learning (Al-Hoorie & Albijadi, 2025). Because how a learner manages uncertainty is itself an aspect of self-regulation, it can complicate the correlation between SRL and vocabulary acquisition differently across frequency bands.

Lexical size also limits the range of strategies available to learners. Those with smaller vocabularies often struggle with more elaborate strategies, such as keyword mnemonics, semantic mapping, and writing example sentences that create rich connections to new words; below a certain vocabulary threshold, these more demanding strategies give way to simple rote rehearsal (Fatemipour & Najafgholikhah, 2015). Band-level scores are also only as reliable as the number of items on which they are based; therefore, drawing strong conclusions from a single band score on a short test can only be tentative.

In our design, we selected items from JACET Levels 2 to 6 and examined whether SRLvoc relates differently to performance at each level, analyzing form and meaning recognition separately. The reasons for sampling these levels are given in Section 3.2.2.

2.5. PROFICIENCY AS A MODERATOR

Proficiency is associated with both how learners regulate their learning (Habók et al., 2022) and the extent to which that regulation is effective (Deng & Trainin, 2023). Among Japanese university students, those with higher proficiency tend to engage more in metacognitive regulation, maintain effort more consistently, and use a broader range of strategies than their less proficient peers (Fukuda, 2018; Ueno & Takeuchi, 2023). This difference in regulatory habits helps explain the proficiency gap. Less proficient learners more frequently report exam-driven motivation and employ a narrower set of vocabulary learning strategies, which can limit the effectiveness of self-regulation in vocabulary acquisition (Boroughani et al., 2023). Intervention studies support these findings, showing that improvements in regulation occur more slowly among weaker learners (Deng & Trainin, 2023). Much of this evidence concerns self-regulation in language learning generally rather than SRLvoc specifically; we treat it as grounds for expecting proficiency to moderate the vocabulary-specific relationship examined here.

Some of this can be attributed to resources. Multivariate studies indicate that proficiency, regulation, and cognitive factors work together rather than independently (Sun et al., 2023). Experiments on scheduling and retrieval likewise show that both the spacing of study sessions and the difficulty of retrieval contribute to long-term retention (Duplice, 2025; Nakata, 2015). A learner requires a certain level of lexical and cognitive foundation before good intentions translate into effective practice; without this foundation, the same amount of effort yields diminished returns (e.g., Fatemipour & Najafgholikhah, 2015; Sun et al., 2023). Therefore, we consider proficiency a likely moderator of the SRLvoc–vocabulary relationship.

Work on this topic has primarily relied on separate regressions or simple high–low comparisons (e.g., Fukuda, 2018; Ueno & Takeuchi, 2023). Multi-group structural equation modeling (SEM), which can directly assess whether the structural relationships among SRL, vocabulary, and proficiency differ between groups, has, to our knowledge, not been applied to this question in Japanese university EFL settings. We conducted such an analysis, comparing lower- and higher-proficiency learners.

2.6. METHODOLOGICAL IMPLICATIONS AND RESEARCH GAP

Three gaps emerge from this review. First, measures of self-regulation and vocabulary are often only loosely aligned: general SRL scales not designed for vocabulary are paired with narrow vocabulary tests, and vocabulary is then reduced to a single score that obscures recognition direction and frequency band. Second, although recent research treats the SRCvoc and related measures as multidimensional, few SEM studies of vocabulary have modeled SRL in this way

or connected it to multiple outcomes. The third gap, which we emphasize most strongly, is that lexical level and proficiency are rarely modeled together, with proficiency indexed by a separate, external measure. This is partly because vocabulary size is so routinely used as a proxy for proficiency that the two constructs collapse into one, rather than being treated as related but distinct (Akase, 2021; McDonald & Asaba, 2015; Ueno & Takeuchi, 2023). Our study addresses all three gaps by incorporating a multidimensional SRLvoc factor; two recognition-based vocabulary factors derived from JACET Levels 2 to 6; TOEFL ITP as an independent proficiency index; and supplementary band-level and multi-group analyses.

2.7. RESEARCH QUESTIONS

In light of these considerations, the study poses four research questions (RQ):

1. To what extent does SRLvoc predict recognition-based L2 vocabulary knowledge among Japanese university students?
2. Does the SRLvoc–vocabulary relationship differ between form recognition and meaning recognition?
3. Do SRLvoc–vocabulary associations vary across JACET 8000 Levels 2–6?
4. Do the relationships among SRLvoc, vocabulary knowledge, and language proficiency differ between lower- and higher-proficiency learners?

RQ1 and RQ2 are addressed with a single-group SEM; RQ3 and RQ4 extend this to individual frequency bands and proficiency subgroups, where the smaller per-band item counts and the median split call for more cautious interpretation.

Figure 1 presents the main SEM: a single common SRLvoc factor, indicated by its three components, is specified as a predictor of two recognition-based vocabulary factors. Additionally, form recognition is modeled as a direct predictor of meaning recognition. The TOEFL ITP is included as an external proficiency factor. Band-level and proficiency-group analyses were planned as follow-ups.

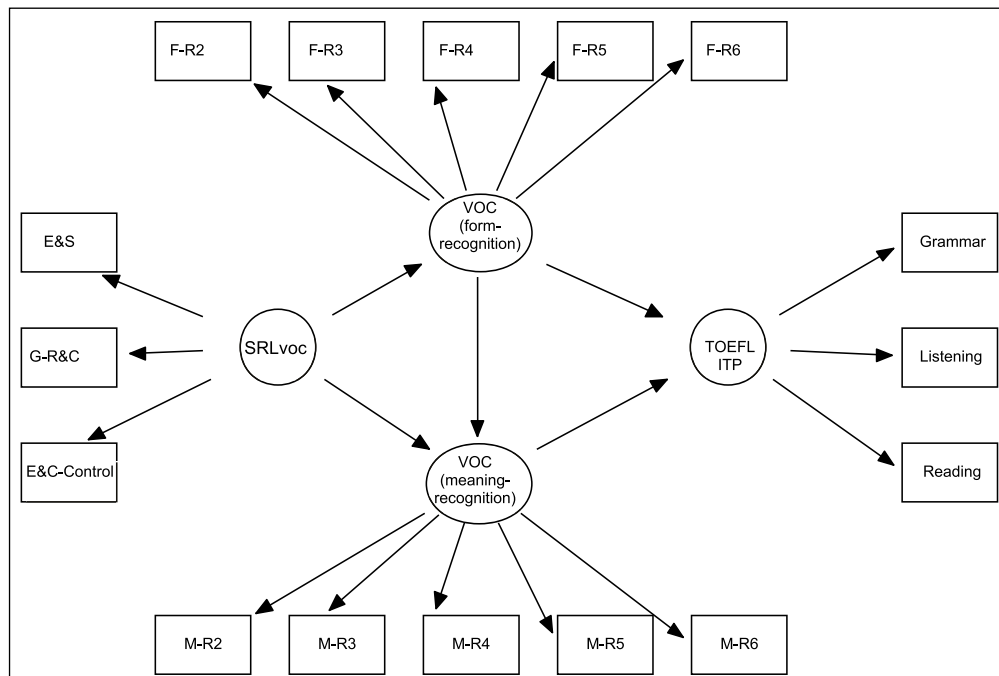


Figure 1 Hypothesized structural model of SRLvoc, form recognition, meaning recognition, and TOEFL ITP proficiency.

Note. VOC = Vocabulary; SRLvoc = Vocabulary-specific self-regulated learning; E&S = Emotion and satiation control; G-R&C = Goal regulation and commitment; E&C-Control = Environmental and cognitive control; F-R2–F-R6 = Level 2–6 form-recognition scores; M-R2–M-R6 = Level 2–6 meaning-recognition scores.

3. METHOD

3.1. PARTICIPANTS

The dataset comprised 267 first-year undergraduates majoring in English at a private university in eastern Japan. Of these, 189 were female (71%) and 78 were male (29%). All students were required to take several English courses, typically attending six to eight classes per week. The TOEFL ITP was administered as part of the university's placement procedures for first-

year English courses. Participation in the vocabulary test and the SRLvoc questionnaire was voluntary; students were informed that their anonymized scores might be used for research purposes and that they could opt out without any disadvantage.

3.2. INSTRUMENTS

3.2.1. L2 proficiency

L2 English proficiency was assessed using the TOEFL ITP, which all first-year students completed before classes began. The university uses this test to place students into proficiency-banded English courses. Scores from the Listening, Grammar, and Reading sections were obtained from all participants. In the SEM analyses, these three section scores were modeled as indicators of a TOEFL proficiency factor; for zero-order correlations (Section 4.1), they were combined into a single composite score.

The test was administered in large lecture halls under standardized conditions approximately two weeks before the semester began. Item-level data were not available from the local ETS office; therefore, internal consistency could not be recalculated. According to the test manual, section reliability for comparable populations ranges from about 0.88 to 0.93 (ETS, 2023). Because the TOEFL ITP targets receptive skills and grammatical knowledge, and does not include speaking or writing components, we treat it as an indicator of receptive and grammatical proficiency (specifically, the listening and reading comprehension and grammatical knowledge assessed by its three sections) rather than as a measure of communicative competence. In this dataset, the TOEFL ITP primarily serves as an external proficiency index against which the SRLvoc–vocabulary relationships can be interpreted.

3.2.2. Vocabulary tests

Vocabulary size was assessed using a test based on the JACET 8000 list. Target words were randomly selected from Levels 2 through 6. Levels 1, 7, and 8 were excluded because pilot testing indicated that Level 1 items were too easy, while Levels 7 and 8 were too difficult for this cohort, leaving the middle bands where these students' knowledge was actively developing.

The 80-item instrument contained two recognition-based sections completed by all participants. In the form-recognition section, a Japanese (L1) prompt was presented, and participants selected the corresponding English (L2) word from four options. In the meaning-recognition section, an L2 prompt was presented, and participants chose the appropriate L1 meaning from four options.

Two practical considerations led us to choose recognition tasks over recall. First, recognition is the skill that most first-year learners are still developing, making it the area where their knowledge varies most and where any SRLvoc signals are likely to be most detectable. Second, recognition allowed us to maintain a consistent test format across all 80 items and all 267 students within a single class session. Incorporating recall tasks would have lengthened the test and introduced challenges related to partial credit and spelling tolerance, which, in our view, would have compromised comparability more than the tasks would have added value.

Each section contained 40 items, distributed across levels as follows: five at Level 2; ten each at Levels 3, 4, and 5; and five at Level 6. More items were allocated to Levels 3–5 to capture the middle of the cohort's range, while keeping the administration of the test to a single class session. Examples are provided in Table 2, with additional sample items in Appendix B.

The two sections were constructed around different target words drawn from the same JACET levels; therefore, we did not assume item-by-item equivalence and interpreted the form–meaning contrast at the level of balanced measures rather than matched pairs. The sections also differed in how their distractors were designed, with implications that we address in the Discussion. In the form-recognition section, the four options were often orthographically similar (e.g., “conform,” “contempt,” “covenant,” and “compensate”), so these items partially tested discrimination among visually similar forms. In contrast, the meaning-recognition section presented options that were semantically unrelated and did not impose a comparable visual-similarity demand. This difference was not entirely consistent within the form section. Moreover, we did not formally equate item difficulty across the two sections (for example, by pretesting and matching items based on facility values), so we consider the difficulty asymmetry between sections a limitation rather than a controlled feature (Sections 5.2 and 6).

FORM RECOGNITION	MEANING RECOGNITION
1. 埋め合わせる、補償する	2. Academic
a) conform	a) 礼儀正しい
b) contempt	b) 正確な
c) covenant	c) 抽象的な
d) compensate	d) 大学の

Table 2 Examples of test items from the form-recognition and meaning-recognition measures.

3.2.3. Pilot analysis of the SRLvoc scale

The SRLvoc scale was developed from the SRCvoc (Tseng et al., 2006) through a pilot analysis, which motivated both the length and the structure of the final instrument. We re-analyzed archived pilot responses to the 20-item Japanese translation of the SRCvoc item pool ($N = 369$). Internal consistency was high ($\alpha = .92$), with corrected item-total correlations ranging from .34 to .77. A principal-axis factor analysis yielded a three-factor solution with initial eigenvalues of 8.10, 1.82, and 1.62; after oblimin rotation, primary factor loadings ranged from .32 to .86. The three factors did not reproduce the five original facets. Tseng et al.'s emotion and satiation controls loaded together and became emotion and satiation control (managing stress or boredom during vocabulary study); commitment and metacognitive controls, both related to setting goals and directing effort, merged into goal regulation and commitment (sustaining effort toward vocabulary goals and resisting procrastination); and environmental control, along with items on sustaining concentration, formed environmental and cognitive control (arranging study conditions and maintaining focus). Items that loaded weakly or cross-loaded were excluded, reducing the translated pool to the 15 items listed in Appendix A.

We adapted rather than reproduced the original instrument for three reasons. First, the original was developed and validated with learners in a different context, so the Japanese translation required re-examination before use. Second, a shorter form was necessary to accommodate the proficiency and vocabulary measures within a single orientation session. Third, as reported above, the pilot analysis supported three stable factors rather than the five original facets. In the resulting framework, the metacognitive, goal-directed function is encompassed within goal regulation, while the cognitive aspects of managing attention and surroundings fall under environmental and cognitive control.

3.2.4. The SRLvoc scale in the main study

In the main study, SRLvoc was assessed with the resulting 15-item scale. Each item was rated on a six-point Likert scale from 1 ("completely disagree") to 6 ("completely agree").

In the main dataset, the three component means were strongly intercorrelated ($r = .72$ to $r = .79$), supporting the use of a composite score. Therefore, for descriptive statistics and zero-order correlations, we report a composite mean across the three components. However, in the SEM analyses, the three component scores serve as indicators of a single common SRLvoc factor, preserving the instrument's multidimensional structure rather than collapsing it into a unidimensional scale.

3.3. PROCEDURE

Data were collected during the placement and orientation period at the start of the academic year under standardized supervision by trained staff in lecture halls or computer labs. Students completed the paper-based TOEFL ITP about two weeks before the spring semester as part of routine placement. During the first week of the semester, a university research center administered the JACET 8000-based vocabulary test in scheduled sessions. As part of a broader survey of learning behaviors, the adapted SRLvoc scale was also administered either immediately after the vocabulary test or in a separate session within the same orientation period. After data collection, all personally identifying information was removed, and each participant was assigned a numeric code.

In the vocabulary test, all students completed the meaning-recognition section (Part 1) before the form-recognition section (Part 2); the order was fixed rather than counterbalanced. Because every student completed the sections in the same order, any practice or fatigue effects are confounded with the section variable, a point we revisit when interpreting the form-meaning contrast (Sections 5.2 and 6).

3.4. DATA ANALYSIS

We first used SPSS 29.0 (IBM Corp., 2022) to compute descriptive statistics, internal consistency coefficients, and Pearson correlations for the adapted SRLvoc, the form-recognition and meaning-recognition vocabulary composites, and the TOEFL ITP composite. Subsequently, SEM was conducted using EQS 6.4 (Bentler & Wu, 2018).

The main SEM addressed RQ1 and RQ2 (Figure 1). SRLvoc was modeled as a latent factor indicated by its three components: emotion and satiation control, goal regulation and commitment, and environmental and cognitive control. Vocabulary knowledge was represented by two recognition-based latent factors (form recognition and meaning recognition), each indicated by five level-specific scores from JACET Levels 2–6. The model also specified a directional path from form recognition to meaning recognition to reflect the assumption that recognizing meanings from L2 forms depends partly on the stability of form-level recognition knowledge. TOEFL proficiency was modeled as a latent factor indicated by listening, grammar, and reading scores. This setup reflects the scope of the study: SRLvoc was expected to relate most directly to the recognition-based vocabulary measures, with TOEFL proficiency providing a broader framework for interpreting those relationships.

To follow up on the main model, we conducted separate regressions in which SRLvoc predicted level-specific proportion-correct scores at JACET Levels 2 through 6 for each recognition measure. Because each level contained only five to ten items, these band-level analyses are considered supplementary. Additionally, we estimated a multi-group SEM to examine whether the SRLvoc-vocabulary-TOEFL relationship differed by proficiency. Students were divided into lower- and higher-proficiency groups based on a median split of the TOEFL composite score.

Model fit was evaluated using the χ^2 statistic, the Comparative Fit Index (CFI), the Non-Normed Fit Index (NNFI), and the Root Mean Square Error of Approximation (RMSEA; Hu & Bentler, 1999). Standardized path coefficients and indirect effects were employed to interpret and compare paths; however, paths involving the TOEFL factor were interpreted cautiously when residual variance approached zero. All SEMs were estimated with maximum likelihood.

4. RESULTS

4.1. DESCRIPTIVE AND RELIABILITY STATISTICS

After screening, complete data for all focal variables were available for 267 students. Table 3 presents descriptive statistics. TOEFL ITP section means fell in the low-to-mid 40s, corresponding to lower-intermediate to intermediate proficiency in this context. Vocabulary scores showed the expected frequency gradient: both measures were near ceiling at Level 2 and dropped into the .60 range by Levels 4–6, with meaning recognition slightly higher throughout, indicating that recognizing meanings from L2 forms was marginally easier than recognizing forms from L1 cues. SRLvoc component means fell in the upper half of the six-point scale, suggesting that most students perceived themselves as at least moderately self-regulated with respect to vocabulary learning.

MEASURE	<i>M</i>	<i>SD</i>	MIN	MAX
TOEFL ITP				
Listening	45.39	3.72	32.00	57.00
Grammar	40.60	5.05	31.00	56.00
Reading	43.85	4.58	31.00	55.00
Vocabulary: Form recognition (proportion correct)				
Level 2	.96	.09	.60	1.00
Level 3	.79	.15	.40	1.00
Level 4	.61	.16	.20	1.00
Level 5	.60	.16	.10	1.00
Level 6	.63	.22	.00	1.00

(Contd.)

Table 3 Descriptive statistics for main study variables (*N* = 267).

Note. Vocabulary level scores are proportions correct (0–1); composite scores are the mean of the five level-specific proportions, also on a 0–1 scale. Internal consistency (Cronbach's α): form recognition = .73, meaning recognition = .72; emotion and satiation control = .87, goal regulation and commitment = .82, environmental and cognitive control = .77, SRLvoc composite = .90.

MEASURE	M	SD	MIN	MAX
Form-recognition composite (mean of Levels 2–6)	.72	.11	.46	.96
Vocabulary: Meaning recognition (proportion correct)				
Level 2	.97	.08	.40	1.00
Level 3	.90	.10	.50	1.00
Level 4	.65	.17	.10	1.00
Level 5	.61	.18	.10	1.00
Level 6	.66	.21	.00	1.00
Meaning-recognition composite (mean of Levels 2–6)	.76	.10	.42	.96
SRLvoc (1–6 scale)				
Emotion and satiation control	4.00	1.09	1.00	6.00
Goal regulation and commitment	3.85	1.00	1.00	6.00
Environmental and cognitive control	4.40	0.85	1.00	6.00
SRLvoc composite (mean of three components)	4.08	0.90	1.08	6.00

Internal consistency was acceptable for all measures (Table 3 note), with the SRLvoc composite at .90. TOEFL ITP section reliability for similar populations falls in the high .80s to low .90s (ETS, 2023), so the measures were sufficiently reliable for the planned analyses.

As shown in Table 4, SRLvoc exhibited a weak but statistically significant correlation with form-recognition vocabulary ($r = .18, p = .004$) and a stronger correlation with meaning-recognition vocabulary ($r = .41, p < .001$). The two recognition measures were themselves strongly correlated ($r = .58, p < .001$). The SRLvoc–TOEFL correlation was not statistically significant ($r = .10, p = .100$), whereas both form recognition ($r = .34$) and meaning recognition ($r = .33$) showed moderate correlations with TOEFL (both $ps < .001$). In summary, SRLvoc was more clearly associated with recognition-based vocabulary than with the broader TOEFL proficiency measure; this SRLvoc–proficiency dissociation is revisited in Section 5.4.

VARIABLE	1	2	3	4
1. SRLvoc (total)	1.00	–	–	–
2. Vocabulary: Form recognition (composite)	.18**	1.00	–	–
3. Vocabulary: Meaning recognition (composite)	.41**	.58**	1.00	–
4. TOEFL ITP total score	.10	.34**	.33**	1.00

Because the composite measure could obscure differences among the three components, we also examined how each component related to the two vocabulary composites. The component-level zero-order correlations revealed a consistent pattern. Emotion and satiation control correlated weakly with form recognition ($r = .17, p = .006$) and moderately with meaning recognition ($r = .39, p < .001$). Goal regulation and commitment exhibited the same pattern, showing a weak correlation with form recognition ($r = .14, p = .026$) and a moderate correlation with meaning recognition ($r = .36, p < .001$). Environmental and cognitive control also correlated weakly with form recognition ($r = .19, p = .002$) and moderately with meaning recognition ($r = .38, p < .001$). Thus, all three components were more closely aligned with meaning recognition than with form recognition, with emotion and satiation control demonstrating the numerically strongest association with meaning recognition, although the three meaning-recognition correlations were similar in magnitude.

4.2. SRLvoc AS A PREDICTOR OF OVERALL RECOGNITION-BASED VOCABULARY KNOWLEDGE

To address RQ1, we estimated the single-group SEM depicted in Figure 2. The model demonstrated good fit: $\chi^2(97) = 130.48, p = .013$, CFI = 0.97, NNFI = 0.97, RMSEA = 0.04, 90% CI [0.02, 0.05]. SRLvoc, modeled as a latent factor with three indicators, showed a significant positive standardized path to form recognition ($\beta = .21$) and a positive but nonsignificant standardized path to meaning recognition ($\beta = .36$). Form recognition showed a large but nonsignificant

Table 4 Correlations among SRLvoc, vocabulary composites, and TOEFL ITP ($N = 267$).

Note. Pearson correlations. SRLvoc (total) = mean of the three component scores; vocabulary composites = means of the five level-specific proportions; TOEFL ITP total = composite of the Listening, Grammar, and Reading sections. ** $p < .01$. The SRLvoc–TOEFL correlation was not significant ($p = .100$).

standardized path to meaning recognition ($\beta = .81$). The squared multiple correlations were .04 for the latent form-recognition factor and .90 for the latent meaning-recognition factor. Inferential support for the SRLvoc-vocabulary relationship therefore comes primarily from the zero-order correlations reported above: SRLvoc correlated with form recognition at $r = .18$ and with meaning recognition at $r = .41$, $p = .004$ and $p < .001$, respectively. The nonsignificant meaning-recognition paths reflect estimation instability from the strong overlap between the two recognition factors rather than evidence of no association.

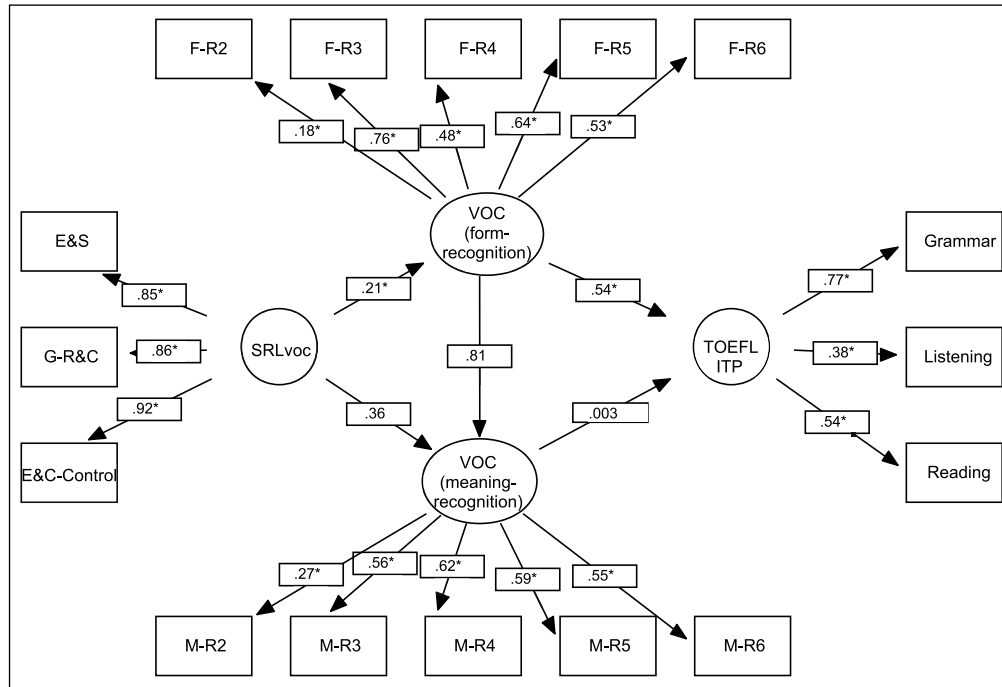


Figure 2 Structural model of SRLvoc, form recognition, meaning recognition, and TOEFL ITP proficiency (standardized solution, full sample).

Note. Coefficients are standardized. Asterisks indicate $p < .05$. VOC = Vocabulary; SRLvoc = Vocabulary-specific self-regulated learning; E&S = Emotion and satiation control; G-R&C = Goal regulation and commitment; E&C-Control = Environmental and cognitive control. The path from meaning recognition to TOEFL is reported to three decimals because it rounds to .00 at two decimals.

In simple terms, the model conceptualizes SRLvoc as a single underlying tendency manifested through its three components. Form recognition and meaning recognition are modeled as two interrelated vocabulary recognition abilities, each estimated from performance across five frequency levels. The SEM provides statistically reliable support for the path from SRLvoc to form recognition; the positive coefficient for meaning recognition is interpreted descriptively and is consistent with the zero-order correlation reported above.

In this cohort, the overlap between recognition vocabulary and TOEFL is primarily driven by form recognition: once both recognition factors are included in the model, meaning recognition adds essentially no direct contribution to TOEFL beyond what form recognition already accounts for. Because the zero-order SRLvoc-TOEFL correlation was nonsignificant and the TOEFL factor exhibited little residual variance, the TOEFL paths are interpreted descriptively (as reflecting overlap between recognition vocabulary and the TOEFL measure) rather than as evidence of an SRLvoc effect on broad proficiency.

4.3. FORM VERSUS MEANING RECOGNITION

RQ2 examined whether SRLvoc was differentially related to the two recognition directions. SRLvoc's association with meaning recognition ($r = .41$) was significantly stronger than its association with form recognition ($r = .18$); a Williams test for two dependent, overlapping correlations confirmed the difference, $t(264) = 4.47$, $p < .001$. The SEM coefficients were descriptively consistent with this ($\beta = .36$ vs. $\beta = .21$), though the meaning-recognition path was not statistically reliable given the overlap between the two recognition factors. This pattern should be interpreted alongside the large estimated coefficient for the path from form recognition to meaning recognition ($\beta = .81$), which indicates substantial overlap between the two recognition abilities but did not reach significance in the full-sample SEM. Thus, the form-meaning difference should not be read as two fully independent SRLvoc effects. The results indicate that SRLvoc is more strongly associated with recognizing meanings from L2 forms than with recognizing L2 forms from L1 cues.

4.4. SUPPLEMENTARY ANALYSES ACROSS JACET FREQUENCY BANDS

RQ3 examined whether the SRLvoc–vocabulary link varied across JACET Levels 2 through 6. For form recognition, SRLvoc did not predict Level 2 performance (very high-frequency items; $B = -.002$, $p = .768$, $R^2 = .00$). A small but significant effect emerged at Level 3 ($B = 0.024$, $p = .021$, $R^2 = .02$). No significant associations were found at Levels 4 or 5 ($p = .217$ and $p = .241$, respectively). Level 6 was again significant ($B = 0.054$, $p < .001$, $R^2 = .05$). Overall, the form-recognition effects were weak and inconsistent across levels.

For meaning recognition, the pattern was somewhat clearer. There was no significant effect at Level 2 ($B = 0.007$, $p = .213$, $R^2 = .01$), but from Level 3 onward, SRLvoc predicted performance: Level 3 ($B = 0.017$, $p = .012$, $R^2 = .02$); Level 4 ($B = 0.090$, $p < .001$, $R^2 = .23$); Level 5 ($B = 0.066$, $p < .001$, $R^2 = .11$); and Level 6 ($B = 0.043$, $p = .003$, $R^2 = .03$).

Interpreted with caution, these follow-up analyses suggest that the highest-frequency items left little room for SRLvoc-related variance, while several mid-range meaning-recognition bands were somewhat more sensitive.

4.5. PROFICIENCY-GROUP DIFFERENCES IN SRLvoc–VOCABULARY–TOEFL RELATIONS

RQ4 examined whether the structural relationships among SRLvoc, vocabulary, and TOEFL proficiency differed between lower- and higher-proficiency learners. We estimated a multi-group SEM with lower- ($n = 134$) and higher-proficiency ($n = 133$) groups, with the same four latent variables in each group: TOEFL proficiency (listening, grammar, and reading), form recognition (Levels 2–6), meaning recognition (Levels 2–6), and SRLvoc (three components).

Model fit was good and consistent across groups: higher-proficiency, $\chi^2(97) = 116.07$, $p = .091$, CFI = 0.96, NNFI = 0.96, RMSEA = 0.04, 90% CI [0.00, 0.06]; lower-proficiency, $\chi^2(97) = 114.63$, $p = .107$, CFI = 0.97, NNFI = 0.96, RMSEA = 0.04, 90% CI [0.00, 0.06]. In both groups, the residual variance of the TOEFL factor was at or near zero; therefore, paths involving TOEFL should be interpreted with caution.

In the higher-proficiency group, SRLvoc had a small, nonsignificant direct effect on form recognition ($\beta = .24$) and a moderate, nonsignificant effect on meaning recognition ($\beta = .40$). The path from form recognition to meaning recognition was large ($\beta = .67$). In the lower-proficiency group, SRLvoc again showed only a small, nonsignificant direct effect on form recognition ($\beta = .13$), but a moderate effect on meaning recognition that reached significance ($\beta = .36$). The path from form recognition to meaning recognition was larger still ($\beta = .89$). Because the vocabulary-to-TOEFL paths were unstable due to near-zero TOEFL residual variance, we do not draw conclusions regarding their exact magnitude or direction. The same latent structure fit both groups; the path from SRLvoc to meaning recognition reached significance only in the lower-proficiency group, although the two estimates were similar in size ($\beta = .36$ in the lower and $\beta = .40$ in the higher group).

5. DISCUSSION

5.1. SRLvoc AS A PREDICTOR OF VOCABULARY KNOWLEDGE

For RQ1, SRLvoc showed a small-to-moderate but consistent association with recognition-based vocabulary knowledge. In the zero-order correlations, SRLvoc was more strongly associated with meaning recognition ($r = .41$) than with form recognition ($r = .18$). In the SEM, SRLvoc showed a significant positive path to form recognition ($\beta = .21$) and a positive but nonsignificant path to meaning recognition ($\beta = .36$). The model also estimated a large but nonsignificant path from form recognition to meaning recognition ($\beta = .81$), indicating that the meaning-recognition factor should be interpreted as closely linked to learners' form-recognition knowledge. Overall, the evidence supports an association between SRLvoc and recognition vocabulary, but the stronger link to meaning recognition rests primarily on the zero-order correlations and their direct comparison (Section 4.3) rather than on a statistically reliable SEM path.

Two factors limit the claim. First, both outcomes were measured using multiple-choice recognition tests, so the findings pertain only to recognition, not to productive retrieval or

depth of knowledge. Second, the effect sizes are partly constrained by the test design: when part of a test approaches ceiling performance, as with the highest-frequency items, there is little variance left for SRLvoc to explain. Within this context, the results support frameworks that conceptualize the regulation of motivation, emotion, and study conditions as a general resource for sustaining vocabulary learning (Li & Wang, 2024; Tseng et al., 2006).

Because we modeled SRLvoc as a single common factor, the present analyses do not isolate the unique contributions of its three components, which were in any case too intercorrelated ($r = .72$ to $r = .79$) to enter separately without collinearity. The component-level correlations reported in Section 4.1 followed the same pattern as the composite: all three related more strongly to meaning than to form recognition. Determining which component accounts for the meaning-recognition association would require designs powered to estimate the components separately; here, the link is best read as a shared regulatory capacity rather than the effect of any one component.

The pedagogical implications require careful consideration. Our data do not demonstrate that an “integrated” delivery of SRLvoc outperforms a standalone approach, as we did not directly compare the two. The data support a narrower point: the association we observed concerns recognition knowledge of the kind built through repeated exposure and review as part of regular study. This pattern suggests that regulation is more helpful when it is connected to the vocabulary work students are already engaged in (e.g., weekly word targets, brief spaced-review sessions, short study logs, and moments of reflection triggered by boredom or distraction) than when taught as a separate skill unrelated to specific words. By “integrated,” we mean embedding these prompts within existing vocabulary tasks; by “add-on,” we refer to a detached self-regulation unit. This distinction is a hypothesis inspired by our results, not a conclusion they confirm, and we present it accordingly.

5.2. RECOGNITION-MEASURE DIFFERENCES IN THE SRLvoc–VOCABULARY RELATIONSHIP

The stronger association with meaning recognition (RQ2) should be read narrowly. Because both outcomes were recognition measures and the two factors overlapped substantially (form-recognition to meaning-recognition, $\beta = .81$), the difference does not reflect two independent SRLvoc effects but rather a stronger link to one type of supported recognition than the other. This interpretation is limited to recognition-based multiple-choice measures: form recognition may depend more on stable orthographic encoding and deliberate L1-to-L2 retrieval, which a broad SRLvoc composite only approximates.

Two features of the test temper this interpretation, both noted earlier (Sections 3.2.2 and 3.3). Because the form options were often orthographic look-alikes while the meaning options were semantically unrelated, part of the weaker SRLvoc–form association and its lower explained variance may reflect how the form options were written rather than a property of form knowledge. Further, because every student took the form section second, in a fixed order, any practice or fatigue effect is confounded with section, which could further widen the gap. Neither can be ruled out here, so we read the form-meaning difference as suggestive rather than conclusive.

In relation to vocabulary teaching, these results favor a staged approach over treating the two recognition directions as interchangeable. Meaning recognition can be supported through extensive reading, quick post-reading meaning checks, and spaced review of previously encountered words. To advance that knowledge toward form access, reasonable additions include short L1-to-L2 retrieval quizzes, cumulative flashcard reviews, and brief reflections on strategies that aided concentration during retrieval.

5.3. PATTERNS ACROSS JACET FREQUENCY BANDS

Because each JACET band contained only five to ten items, these patterns are read with caution. As reported in Section 4.4, the near-ceiling Level 2 items left little room for SRLvoc variance, whereas several mid-range meaning bands were more sensitive. The primary takeaway is methodological: reporting results by level can reveal heterogeneity that a single composite measure conceals. However, stronger claims about mid-frequency words being a particular locus of SRLvoc influence would require denser sampling. Therefore, we do not

argue that SRLvoc instruction should target a specific band. A more defensible position is that regulation support should be integrated throughout the vocabulary curriculum. Future research employing more items per band, item-response modeling, or matched word sets is needed to determine whether any band benefits disproportionately.

5.4. PROFICIENCY DIFFERENCES IN SRLvoc–VOCABULARY ASSOCIATIONS

RQ4 examined whether the pattern differs by proficiency. Because the groups came from a median split and the TOEFL factor showed boundary estimates, the comparison is exploratory and simplifies what is really a continuous scale. With this caveat, the SRLvoc-meaning recognition estimates were comparable across groups ($\beta = .40$ in the higher- and $\beta = .36$ in the lower-proficiency group) and reached significance only in the lower group; because we did not test the cross-group difference directly, we do not read this as moderation by proficiency. The most cautious interpretation is that self-regulatory routines are more apparent while learners are still developing their recognition lexicon; however, the study design does not support a developmental claim.

One result warrants comment because it contradicted expectations. SRLvoc and TOEFL scores were essentially uncorrelated ($r = .10$, $p = .100$), despite both relating to vocabulary. We had anticipated a positive, albeit weak, association based on the reasoning that better-regulated learners should accumulate more of the receptive knowledge assessed by the TOEFL. This null finding aligns with a recurring pattern in the literature, where self-reported regulation and strategy use correlate only weakly and inconsistently with global proficiency measures (Fukuda, 2018; Ueno & Park, 2025). Proficiency tests draw on listening, reading, and grammar skills developed over years of varied input, whereas a vocabulary-specific SRL scale captures current study habits; thus, the two need not move in tandem. In our model, the TOEFL factor retained almost no residual variance once vocabulary was included, so the vocabulary-to-TOEFL paths mainly register this shared variance rather than any effect of SRLvoc on broad proficiency. The broader implication is that SRL, vocabulary, and proficiency are related but distinct constructs: SRLvoc tracked vocabulary, and vocabulary tracked TOEFL, but SRLvoc was not directly related to TOEFL performance.

6. CONCLUSION

This study examined how SRLvoc relates to L2 vocabulary knowledge among Japanese university EFL learners, treating vocabulary as recognition knowledge differentiated by retrieval direction, frequency band, and learner proficiency. Using a three-component SRLvoc measure, a JACET 8000-based test that distinguishes form recognition from meaning recognition, and TOEFL ITP scores, we found that SRLvoc showed a small-to-moderate but consistent association with recognition vocabulary, clearer for meaning recognition than for form recognition, with comparable estimates across proficiency groups. The band-level analyses suggested that the highest frequency items left little room for SRLvoc-related variance, and the sparse per-band sampling renders any frequency-specific conclusions tentative.

Several limitations constrain these conclusions: the data were drawn from a single cohort at one institution; the SRLvoc measure was self-reported and captured broad regulatory capacities rather than observed vocabulary-study behavior; the test assessed recognition only; the two sections employed different vocabulary and differently constructed distractors and were administered in a fixed order; item-level TOEFL data were unavailable; and the cross-sectional design precludes causal or developmental inferences.

These limitations suggest a concrete research agenda. Longitudinal designs could determine whether early SRLvoc predicts subsequent vocabulary growth, rather than merely co-occurring with it. Measures of productive vocabulary and depth of knowledge would assess whether the advantage in meaning recognition extends beyond simple recognition. Using item-matched sections with counterbalanced order and pretested difficulty would help distinguish genuine form-meaning differences from artifacts of test construction. Denser sampling within frequency bands, analyzed through item-response models, would enable more accurate claims about specific frequency levels. Estimating the three components of the adapted SRLvoc separately, in sufficiently large samples to avoid the collinearity encountered here, would clarify which aspects of regulation are associated with meaning recognition and refine pedagogical

recommendations. Finally, multi-institutional samples employing a continuous measure of proficiency, rather than a median split, would test the generalizability of the proficiency-group patterns.

The contribution is not that SRLvoc determines broad proficiency or governs specific frequency bands. Rather, SRLvoc becomes visible and interpretable when vocabulary knowledge is treated as recognition knowledge differentiated by retrieval direction rather than as a single undifferentiated score. Its clearest association in this cohort was with recognizing meaning from L2 forms.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Appendices.** Appendices A and B. DOI: <https://doi.org/10.22599/jesla.164.s1>

ETHICS AND CONSENT

The data were collected as part of the university's routine placement and orientation procedures and were anonymized before analysis: all personally identifying information was removed, and each participant was assigned a numeric code. Students were informed that their anonymized scores might be used for research purposes; participation in the vocabulary test and the SRLvoc questionnaire was voluntary, and students could opt out without disadvantage.

AUTHOR AFFILIATIONS

Megumi Sugita  orcid.org/0009-0001-3357-2078
Kanda University of International Studies, Japan

Siwon Park  orcid.org/0009-0004-5561-4162
Juntendo University, Japan

REFERENCES

- Akase, M.** (2021). A longitudinal analysis of growth in vocabulary size and strategy use: Focusing on Japanese EFL senior high school students. *Journal of Pan-Pacific Association of Applied Linguistics*, 25(2), 105–130. <https://doi.org/10.25256/paal.25.2.6>
- Al-Hoorie, A. H., & Albijadi, O.** (2025). The motivation of uncertainty: Gamifying vocabulary learning. *RELC Journal*, 56(2), 332–345. <https://doi.org/10.1177/00336882241229244>
- Alamer, A., Teng, M. F., & Mizumoto, A.** (2024). Revisiting the construct validity of Self-Regulating Capacity in Vocabulary Learning Scale: The confirmatory composite analysis (CCA) approach. *Applied Linguistics*, 45(3), 512–534. <https://doi.org/10.1093/applin/amae023>
- Bentler, P. M., & Wu, E. J. C.** (2018). *EQS 6.4 for Windows* [Computer software]. Multivariate Software, Inc.
- Boroughani, T., Xodabande, I., & Karimpour, S.** (2023). Self-regulated learning with mobile devices for university students: Exploring the impacts on academic vocabulary development. *Discover Education*, 2, Article 5. <https://doi.org/10.1007/s44217-023-00028-z>
- Chen, C., Jamiat, N., & Du, S.** (2024). Effect of a self-regulated vocabulary learning mobile application on EFL students' vocabulary learning achievement and motivation. *International Journal of Interactive Mobile Technologies*, 18(16), 4–23. <https://doi.org/10.3991/ijim.v18i16.49755>
- Deng, Q., & Trainin, G.** (2023). The effect of a self-regulated intervention on vocabulary knowledge and self-regulated learning skills for English language learners. *Reading Psychology*, 44(4), 357–383. <https://doi.org/10.1080/02702711.2023.2187908>
- Duplice, J.** (2025). Generation and L2 vocabulary learning: A classroom action study on the efficacy of the generation desirable difficulty in learning L2 vocabulary. *Vocabulary Learning and Instruction*, 14(1), Article 102482. <https://doi.org/10.29140/vli.v14n1.102482>
- ETS.** (2023). *TOEFL ITP: Test and data summary*. <https://www.ets.org/pdfs/toefl/toefl-ityp-test-score-data-2022.pdf>
- Fatemipour, H., & Najafgholikhani, M.** (2015). The impact of self-regulated strategy development on vocabulary learning among English language learners. *Mediterranean Journal of Social Sciences*, 6(5, Suppl. 2), 249–258. <https://doi.org/10.5901/mjss.2015.v6n5s2p249>
- Fukuda, A.** (2018). The Japanese EFL learners' self-regulated language learning and proficiency. *Journal of Pan-Pacific Association of Applied Linguistics*, 22(1), 65–87. <https://doi.org/10.25256/paal.22.1.4>
- Habók, A., Magyar, A., & Molnár, G.** (2022). English as a foreign language learners' strategy awareness across proficiency levels from the perspective of self-regulated learning metafactors. *Frontiers in Psychology*, 13, Article 1019561. <https://doi.org/10.3389/fpsyg.2022.1019561>

- Hashimoto, B. J. (2021). Is frequency enough?: The frequency model in vocabulary size testing. *Language Assessment Quarterly*, 18(2), 171–187. <https://doi.org/10.1080/15434303.2020.1860058>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- IBM Corp. (2022). *IBM SPSS Statistics for Windows* (Version 29.0) [Computer software].
- Kudo, M., Mizumoto, A., & Kumazawa, T. (2015). Validation of a vocabulary learning strategy scale and its relationship to vocabulary level test scores. *Asian Journal of English Language Teaching*, 25(1), 81–112. <https://doi.org/10.65961/AJELT-2015-1-004>
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399–436. <https://doi.org/10.1111/j.0023-8333.2004.00260.x>
- Li, J., & Wang, C. (2024). Mediating roles of motivational beliefs and vocabulary learning strategies for the relationship between self-regulation and vocabulary proficiency. *European Journal of Education*, 59(3), Article e12706. <https://doi.org/10.1111/ejed.12706>
- McDonald, K., & Asaba, M. (2015). “I don’t know” use and guessing on the bilingual Japanese vocabulary size test: A preliminary report. *Vocabulary Learning and Instruction*, 4(1), 16–25. <https://doi.org/10.7820/vli.v04.1.mcdonald.asaba>
- Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge University Press.
- Nakata, T. (2015). Effects of expanding and equal spacing on second language vocabulary learning: Does gradually increasing spacing increase vocabulary learning? *Studies in Second Language Acquisition*, 37(4), 677–711. <https://doi.org/10.1017/S0272263114000825>
- Nakata, T. (2016). Effects of retrieval formats on second language vocabulary learning. *International Review of Applied Linguistics in Language Teaching*, 54(3), 257–289. <https://doi.org/10.1515/iral-2015-0022>
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.
- Oxford, R. (2016). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315719146>
- Park, S., Fukuda, T., & Wadden, P. (2026). The mediating role of vocabulary in linking L1 reading and L2 proficiency. *International Journal of TESOL Studies*, 8(3), 95–112. <https://doi.org/10.58304/ijts.251107>
- Park, S., Ito, Y., & Sugita, M. (2020). Effects of test format in assessing L2 vocabulary knowledge and skills. *Studies in Linguistics and Language Teaching*, 31, 111–124. <https://doi.org/10.69236/0000001843>
- Sasao, Y. (2008). Estimating vocabulary size: Does test format make a difference? *JACET Journal*, 46, 63–76.
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan. <https://doi.org/10.1057/9780230293977>
- Sevigny, P., Mack, L., Stilp, L., & Berger, M. (2024). High-frequency vocabulary: Moving from recognition to recall level on Quizlet. *SAGE Open*, 14(2), 1–12. <https://doi.org/10.1177/21582440241242604>
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *The Language Learning Journal*, 36(2), 139–152. <https://doi.org/10.1080/09571730802389975>
- Sun, D., Chen, Z., & Zhu, S. (2023). What affects second language vocabulary learning? Evidence from multivariate analysis. *Frontiers in Education*, 8, 1210640. <https://doi.org/10.3389/feduc.2023.1210640>
- Tseng, W.-T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, 27(1), 78–102. <https://doi.org/10.1093/applin/ami046>
- Uchiyama, T., Eguchi, M., & Clenton, J. (2022). The contribution of guessing from context and dictionary use to receptive and productive vocabulary knowledge: A structural equation modeling approach. *Language Teaching Research*, 26(6), 1193–1217. <https://doi.org/10.1177/13621688221122138>
- Ueno, S., & Park, S. (2025). Exploring the relationship between self-regulation, grit, motivation, and English proficiency among university students in Japan. *Studies in Self-Access Learning Journal*, 16(2), 418–436. <https://doi.org/10.37237/160208>
- Ueno, S., & Takeuchi, O. (2023). Which vocabulary learning strategies are important and useful for Japanese university students? A text-mining approach. *Vocabulary Learning and Instruction*, 12(1), 18–30. <https://doi.org/10.29140/vli.v12n1.1213>
- Ueno, S., & Takeuchi, O. (2025). Relationship between second language vocabulary knowledge and vocabulary learning strategy use: A meta-analysis of correlational studies. *International Review of Applied Linguistics in Language Teaching*, 63(2), 1115–1137. <https://doi.org/10.1515/iral-2023-0100>
- Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, 30(1), 79–95. <https://doi.org/10.1017/S0272263108080042>
- Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, 26(4), 696–725. <https://doi.org/10.1177/1362168820913998>

TO CITE THIS ARTICLE:

Sugita, M., & Park, S. (2026). Self-Regulated Learning and L2 Vocabulary Knowledge in Japanese University EFL Learners: Form Recognition, Meaning Recognition, and Proficiency-Level Differences. *Journal of the European Second Language Association*, 10(1), 131–146. DOI: <https://doi.org/10.22599/jesla.164>

Submitted: 02 December 2025

Accepted: 03 September 2026

Published: 23 October 2026

COPYRIGHT:

© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>.

Journal of the European Second Language Association, is a peer-reviewed open access journal published by White Rose University Press.

