JESLA Journal of the European Second Language Association

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

RESEARCH

# An investigation of vocabulary development and prosocial peer learning support engagement using Growth Curve Modeling

TETSUYA FUKUDA
PAUL WADDEN
SIWON PARK
MARCEL VAN AMELSVOORT

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

WHITE ROSE UNIVERSITY PRESS
Universities of Leeds, Sheffield & York

## ABSTRACT

This study explored the relationship between vocabulary development and engagement among first-year university students, employing Growth Curve Modeling (GCM). The study examined how prosocial peer learning support (PPLS) engagement predicts vocabulary acquisition, using the New General Service List Test (NGSLT) based on the New General Service List (Browne, 2014) and the PPLS engagement questionnaire (Fukada et al., 2017) as assessment tools. Three different English-English versions (A, C, and B) of the NGSLT were administered at the beginning, middle, and end of an academic year to measure students' vocabulary growth. Given that the NGSLT is a reliable means of tracking vocabulary acquisition over time, GCM was used for statistical analysis. Furthermore, the study explored whether initial vocabulary proficiency and engagement influenced subsequent learning outcomes. The application of GCM provided insights into individual variations while controlling for initial vocabulary knowledge and group composition. Based on previous studies, it was expected that students with stronger engagement in English learning would exhibit steeper vocabulary growth trajectories. The results, however, suggest that PPLS engagement did not meaningfully contribute to vocabulary acquisition. These findings highlight the need to examine the roles of different components of engagement in vocabulary development more carefully and identify potential mediating factors that influence the relationship between engagement and learning outcomes.

**CORRESPONDING AUTHOR:**
**Tetsuya Fukuda**
Juntendo University, JP
t.fukuda.wv@juntendo.ac.jp

**KEYWORDS:**
vocabulary acquisition; NGSLT; PPLS engagement; Japanese EFL learners; Growth Curve Modeling

**TO CITE THIS ARTICLE:**
Fukuda, T., Wadden, P., Park, S., & Van Amelsvoort, M. (2026). An investigation of vocabulary development and prosocial peer learning support engagement using Growth Curve Modeling. *Journal of the European Second Language Association*, 10(1), 165–180. DOI: https://doi.org/10.22599/jesla.171

DOI: 10.22599/jesla.171

# 1. INTRODUCTION

Vocabulary knowledge is a core component of L2 proficiency and is central to learners' academic success and communicative competence (Nation & Webb, 2011). Accordingly, considerable research on vocabulary and vocabulary acquisition has focused on identifying essential lexical resources (e.g., Coxhead, 2000), developing reliable instruments to measure vocabulary size (e.g., Nation & Beglar, 2007), and modeling vocabulary development over time (e.g., De Wilde, 2025). At the same time, research in learner psychology within second language acquisition (SLA) has increasingly emphasized the importance of affective and psychological factors, including social and psychological engagement, in shaping language learning outcomes. Despite these parallel lines of inquiry, research integrating vocabulary development, psychological engagement, and longitudinal analytical approaches remains limited (e.g., De Wilde, 2025). In particular, few studies have examined how engagement interacts with vocabulary growth trajectories over extended periods using robust statistical modeling (e.g., Pan et al., 2026). To address this gap, the present study examined university students' vocabulary development as measured by the New General Service List Test (NGSLT), investigated the role of prosocial peer learning support (PPLS) engagement as a psychological factor, and employed Growth Curve Modeling (GCM) to capture dynamic changes over time. The following literature review synthesizes research on NGSLT-based vocabulary measurement, engagement in L2 learning, psychological factors in vocabulary development, and the application of GCM in SLA research, thereby establishing the theoretical and methodological foundation for the current study.

## 1.1. MEASUREMENT OF VOCABULARY WITH THE NGSLT

West (1953) compiled the General Service List (GSL), which contains about 2,000 word families selected from written text. The purpose of the list was to help learners gain maximum communicative competence with limited instructional time. West believed that mastering the frequently used words in the list would yield the greatest return for learners. For decades, the GSL served as a foundation for English language teaching and learning, material development, and research. However, changes in language use, including new commonly used vocabulary items (e.g., "database," "gender," and "genetic") over half a century and the availability of large electronic corpora led to the development of Browne's (2014) New General Service List, or NGSL, which was based on a much larger and more representative corpus of contemporary English. The NGSL revised and updated the original list, offering improved coverage of modern texts while reflecting current usage more accurately. As a result, Browne's NGSL is often considered a more focused and up-to-date alternative to West's GSL for today's learners and educators.

Following the development of the NGSL, Stoeckel and Bennett (2015) constructed the NGSLT to "assess written receptive knowledge of the words on the New General Service List (NGSL)" (p. 1). As of the time of writing, there are three versions available: Test A (one version in English and one in Japanese), Test B (in English and in Japanese), and Test C (in English). Since its inception, the NGSLT has been validated, and its association with other tests has been investigated in several studies. Stoeckel and Bennett (2015) calculated the correlation coefficient between the results of 33 students who took the NGSLT and the Test of English for International Communication (TOEIC). They found the two results were strongly correlated ($r$ = .72, $p$ < .001). Nadasdy et al. (2018) collected data from 98 Japanese first-year students who took the NGSLT in three versions: English-only (EE), English with target words translated into Japanese (EJ), and students' own translations of the target words (TR). The internal consistency of each version was high at .90 (EE), .83 (EJ), and .89 (TR). The correlations between the NGSLT and TOEIC were much weaker at $r$ = .31 (EE) and $r$ = .37 (EJ) than in Stoeckel and Bennett (2015). More recently, Fukuda et al. (2025) examined the association between the NGSLT and Test of English as a Foreign Language Institutional Testing Program (TOEFL ITP) among 260 first-year Japanese university students. Correlation analysis showed a statistically significant, moderate relationship between NGSLT and TOEFL scores ($r$ = .47, $p$ < .001), indicating that the two measures assess related but distinct aspects of English proficiency. Scatter plot analysis revealed an asymmetrical pattern: high TOEFL scores consistently predicted high NGSLT scores, whereas high NGSLT scores did not necessarily predict strong TOEFL performance.

As outlined above, the NGSLT and its association with other tests, such as TOEIC (Stoeckel & Bennett, 2015) and TOEFL (Fukuda et al., 2025), have been examined in different studies.

DOI: 10.22599/jesla.171

However, few studies have yet examined the relationship between the NGSLT and psychological factors, a gap the current study seeks to address.

## 1.2. ENGAGEMENT

Positive attitudes toward an institution—whether a school or a business—are often associated with productive and constructive behavior. In educational contexts, such behavior is commonly referred to as school engagement. Willms (2003) defined school engagement as "the extent to which students identify with and value schooling outcomes and participate in academic and non-academic school activities" (p. 8). School engagement has widely been conceptualized not as a single trait, but as a multidimensional construct. In Alexander et al. (1997), school engagement consisted of two components: behavioral (e.g., effort and participation) and emotional (e.g., interest and belonging) engagement. Later, Fredricks et al. (2004) proposed a three-factor model of school engagement comprising behavioral, emotional, and cognitive (e.g., self-regulation and investment) dimensions. This three-factor model was validated by Wang et al. (2011) employing confirmatory factor analysis (CFA), and it is still widely employed as a framework and measurement of school engagement. It should be noted that Reeve and Tseng (2011) later introduced a four-factor model by adding agentic (e.g., expressing preferences and asking questions) engagement. Moreover, Zhou et al. (2021) proposed a different four-factor model with cognitive, behavioral, emotional, and social dimensions. However, most empirical studies, including Zhang et al. (2023), have employed the three-factor model of Fredricks et al. (2004).

School engagement has been widely regarded as a contributor to academic achievement. Okunuki and Kashimura (2024) conducted a meta-analysis examining the effects of school engagement on academic performance. Their analysis included 18 journal articles and book chapters that investigated the relationship between engagement and academic achievement, with outcome measures ranging from course grades to self-reported achievement. Overall, engagement was found to have a significant positive effect on academic achievement ($r = .23$, $p < .001$). Each of the four engagement components—cognitive, behavioral, emotional, and other—emerged as significant predictors of academic achievement ($r = .26$, .24, .16, and .23, respectively).

The specific type of school engagement investigated in the current study was PPLS engagement, which was incorporated into a model alongside vocabulary test results. Drawing on possible selves theory (Markus & Nurius, 1986), near-peer role modeling (Murphey & Arao, 2001), and socio-dynamic perspectives (Vygotsky, 1978), Murphey et al. (2014) conceptualized "Ideal Classmates" and introduced an intervention called Ideal Classmates Priming to help students visualize effective learners of English. Subsequently, Fukada et al. (2017) developed a scale to measure the extent to which learners enjoy helping their peers learn English and being helped by others to do so.

Following Fukada et al. (2017), the same group of researchers termed the concept measured with the scale as "PPLS engagement" and defined it as "a type of engagement in which students' perceptions of their mutually constructive participation matter to them in their academic achievement" (Fukuda et al., 2021, p. 182). In their review of Fukuda et al. (2021), Gu and Sun (2021) described PPLS engagement as an "innovative approach to learner engagement" and "persuasively endorsed" its effectiveness in combating learners' passivity and disengagement (p. 2). Taken together, these conceptualizations suggest that PPLS engagement captures a socially grounded and interaction-oriented form of learner involvement that extends beyond individual effort. As such, incorporating PPLS engagement into the present model allows for a more nuanced examination of how collaborative classroom dynamics contribute to language learning outcomes.

PPLS engagement may also be relevant to vocabulary development because vocabulary learning can be supported through interaction with peers. From a Vygotskian perspective, learning is mediated through social interaction, and learners may develop knowledge through collaborative processes with others (Vygotsky, 1978). Students who are more actively involved in mutually supportive peer relationships may have greater opportunities to discuss unfamiliar words, share learning strategies, and encourage one another to sustain vocabulary study. Thus, although vocabulary acquisition ultimately requires individual learning and retention, PPLS engagement may facilitate the social and motivational processes that support vocabulary development over time.

DOI: 10.22599/jesla.171

Empirical research on PPLS engagement also provides some support for these proposed mechanisms. For example, Fukada et al. (in press) found that PPLS engagement was strongly associated with larger gains in temporal self-image motivations (past, present in- and out-of-class, and future). While the students generally showed motivational growth over the semester, those in the high PPLS engagement group exhibited the largest improvements (medium-to-large effect sizes), followed by the medium-level group, with the lower-level group showing the smallest gains. Correlational and ANOVA results demonstrated that motivation became increasingly tied to PPLS engagement over time, and qualitative findings revealed reciprocal patterns in that students who gave more support also received more. The authors concluded that PPLS engagement functioned as a key mechanism through which Ideal Classmates Priming enhanced motivation and classroom dynamics.

To date, however, no studies have reported on the effects of PPLS engagement on academic achievement, including L2 vocabulary development. The current study aims to fill the gap.

### 1.3. PSYCHOLOGICAL FACTORS AND VOCABULARY

Despite the extensive body of research on L2 vocabulary acquisition, particularly on the identification of academic words (e.g., Coxhead, 2000), measurement of vocabulary size (e.g., Nation & Beglar, 2007), and requirements of vocabulary for specific activities (e.g., Webb & Rodgers, 2009), comparatively little research has examined motivational and affective psychological factors underlying vocabulary development. One of the few such studies was conducted by Zhang et al. (2023), who examined how learners engage in digital game-based vocabulary learning and how different types of engagement influence English as a Foreign Language (EFL) vocabulary development. The study involved 50 Chinese university students who participated in a specially designed digital vocabulary-learning game. In terms of engagement, the three-factor model proposed by Fredricks et al. (2004) was employed, as in many other empirical studies. The results showed that engagement was not static but evolved over time, and that overall engagement significantly improved both short-term vocabulary learning and long-term retention. At the same time, among the three factors of engagement, only behavioral engagement showed mixed effects. While behavioral engagement was necessary for vocabulary growth, excessive time or effort sometimes led to fatigue and reduced efficiency.

Furusho and Kitagami (2022) investigated how psychological factors, such as motivation, anxiety, and self-efficacy, influence vocabulary learning among Japanese university EFL learners. They found that intrinsic motivation for English learning was strongly associated with positive attitudes toward vocabulary study, whereas extrinsic, reward-oriented motivation was linked to higher anxiety about vocabulary learning. Importantly, vocabulary size showed a moderate positive correlation with vocabulary learning self-efficacy, but not with general English learning motivation. These results suggest that learners' confidence in vocabulary learning, rather than overall motivation, plays a more direct role in vocabulary development.

Baharloo and Ghasemian (2024) investigated how ambiguity tolerance, which means learners' ability to cope with uncertain or unclear input, and L2 reading anxiety influence the performance of Iranian intermediate English learners when encountering unknown vocabulary in reading comprehension tasks. SEM analyses showed that greater ambiguity tolerance was significantly associated with better performance in dealing with unknown words, while L2 reading anxiety was negatively related to performance on unknown vocabulary. These findings indicate that psychological traits can influence vocabulary learning.

Another relevant study is Park et al. (2025), which examined the relationship between L1 reading ability and L2 English proficiency among Japanese university students, focusing on the mediating role of L2 vocabulary. SEM was used to test direct and indirect relationships, and the results showed that L1 reading positively predicted L2 vocabulary, which in turn strongly predicted L2 proficiency, while the direct effect of L1 reading on L2 proficiency was not significant. These findings suggest that L1 reading contributes to L2 learning mainly by facilitating vocabulary growth.

Although the studies that have investigated psychological factors in vocabulary learning, such as Furusho and Kitagami (2022), Baharloo and Ghasemian (2024), and Park et al. (2025), have yielded important insights, few studies in SLA, except for Zhang et al. (2023) and, more recently, Oikawa and Uchihara (2026), have examined the effect of engagement on vocabulary development. This study aims to fill this gap.

### 1.4. EMPLOYMENT OF GROWTH CURVE MODELING (GCM) IN SLA

The data collected in the current study were analyzed using both traditional statistical methods, such as descriptive statistics, internal consistency, and Analysis of Variance (ANOVA), as well as SEM. By definition, SEM "represents relationships among observed and latent variables" (Purpura, 1999), and GCM is a kind of SEM that allows for the estimation of intra-individual patterns of change over time (Curran et al., 2010). According to Kline (2023), the key features of GCM are that (a) a variable is measured on at least three different occasions, (b) the data are structured over time, and (c) repeated measures on the variable serve as indicators of latent growth curve. In other words, GCM is most appropriate when researchers aim to examine the data from three or more points, which better represent reality than those from two data points (Parsons & McCormick, 2024). It should be noted that Latent Growth Curve Modeling (LGCM) and Growth Curve Modeling (GCM) are considered interchangeable in this review, as in many other studies.

Many researchers (e.g., Csizér & Dörnyei, 2005; Plonsky, 2014; Winke, 2014) have argued for the use of SEM in SLA, and, more recently, Zhou et al. (2023) also advocated greater use of GCM. Given the general interest in dynamic change over time among SLA researchers, GCM should be more frequently employed in SLA studies. Recent SLA studies have applied GCM to psychological and linguistic variables, including foreign language anxiety and enjoyment (Elahi Shirvan et al., 2021), motivation and school belonging (Fukuda, 2020), and speaking development and self-efficacy (Li & Zhang, 2023). Collectively, these studies demonstrate the usefulness of GCM for capturing individual differences and nonlinear change over time.

Nonetheless, relatively few studies have used GCM to examine vocabulary development together with psychological variables. Therefore, the current study partly fills this gap by focusing on vocabulary development and psychological constructs using GCM as the analytical framework.

### 1.5. RESEARCH QUESTIONS

Addressing the gaps identified in the preceding review, the present study integrates vocabulary development, PPLS engagement, and longitudinal modeling within a unified analytical framework. It applies GCM to examine both overall developmental trends and individual differences in vocabulary growth over time. The following research questions (RQs) are designed to clarify the relationships among these constructs.

RQ1. How do students' NGSLT scores change across three data collection points over the academic year?

RQ2. Do students differ meaningfully in (a) initial vocabulary level (intercept) and (b) rate of change (slope)?

RQ3. Does students' PPLS engagement predict the vocabulary growth factor?

RQ4. Is initial vocabulary status (intercept) related to subsequent change (slope), once PPLS engagement is in the model?

## 2. METHODS

This study adopted a longitudinal, quantitative research design to investigate the relationship between PPLS engagement and L2 vocabulary development among Japanese university EFL students during their first academic year. Data were collected through a combination of standardized vocabulary assessments at three time points in April, July, and December, and through a validated engagement questionnaire, administered in June, allowing for the examination of both developmental change in vocabulary knowledge and the role of PPLS engagement in that process. The following sections describe the participants and instructional context, the instruments used to measure vocabulary knowledge and engagement, and the statistical procedures employed to analyze the data.

### 2.1. PARTICIPANTS

The data were collected from first-year students at a university in eastern Japan during the 2025 academic year. Of the 251 students who entered the university in April 2025, 218 agreed

DOI: 10.22599/jesla.171

to participate in the study, and 164 provided usable data for the descriptive analyses. For the GCM analysis, 149 of these 164 cases had complete data on all variables entered into EQS (Bentler & Wu, 2018) and were retained as the complete-case analytic sample. All participants were informed that their anonymized test scores would be used for research purposes, and they were given the opportunity to opt out without penalty. The survey was conducted in June, after the students had had time to acclimate to the university and to develop peer relationships, because the PPLS items assess students' perceptions of mutually supportive interactions with classmates. At the same time, because PPLS engagement was measured only once, the resulting score represents students' perceptions at a mid-year point rather than a baseline or longitudinal measure of engagement. Accordingly, PPLS engagement was treated as a time-invariant predictor in the present model, and possible changes in peer relationships or engagement across the academic year could not be examined. The three versions of the NGSLT were administered as follows: Test A in April, shortly after matriculation; Test C in July, at the end of the first semester; and Test B in December, near the end of the first year. The self-reported mean age was 18.32 as of the first data collection point in April. Their self-reported gender was 59 male (36.0%), 105 female (64.0%), with no participants choosing the "Prefer not to say" option in the background survey. The group's mean TOEFL ITP score in July was 447.44, with scores ranging from 350 to 623. The score was slightly lower than the average TOEFL ITP score for Japanese students in 2024 (ETS, 2025). However, given that TOEFL ITP is commonly taken by those attending upper-tier universities, the participants are assumed to have a mid-level of English proficiency among Japanese university students. Of the 164 participants, more than 90% were Japanese and had presumably undergone at least eight years of English education prior to matriculation, with the remaining students mostly Chinese.

All student participants were enrolled in an international liberal arts program with substantial English instruction in the first and second years. Except for a small number of proficient students who were exempted from some courses, students completed nine hours of English classes per week for 30 weeks. Most lessons focused on academic English through a Content and Language Integrated Learning (CLIL) approach, covering six liberal arts areas: health and medicine, zoology, human history, earth science, economics, and sociology. Separate strands also addressed academic skills and TOEFL preparation.

In addition to content-related vocabulary, students studied two decontextualized word lists: 400 selected NGSL words and 1,800 words from the Global Academic Vocabulary list (Wadden et al., 2016). Instructional materials included bilingual in-house texts with example sentences, Quizlet-based online practice, and weekly vocabulary quizzes. Although teaching approaches varied across instructors, common activities included annotating texts with personalized examples, identifying target words and related word forms, comparing known and unknown words with classmates, creating personalized Quizlet decks, and completing online practice activities. Weekly quizzes were administered through Google Forms, and some instructors required a minimum score before students progressed to the next lesson. Across the year, vocabulary generally became more difficult, progressing from higher-frequency, more familiar words to lower-frequency words that students were less likely to know.

## 2.2. INSTRUMENTS

In this study, the influence of engagement on vocabulary development was examined. In this section, the NGSLT vocabulary test and the PPLS engagement questionnaire are explained.

### 2.2.1. NGSLT

The NGSLT is a 100-item test designed to measure a learner's written receptive knowledge of high-frequency English vocabulary. It is based on the 2,809 headwords in the NGSL, which provides approximately 92% coverage of the running words (tokens) found in most general English texts and about 86% coverage of the words in typical academic texts. The 100 multiple-choice items are divided equally among five frequency bands of the NGSL, each comprising approximately 560 words. By targeting specific bands of words, the test provides precise information about which frequency levels a learner has mastered (typically a score of 80–85% or higher), allowing targeted study and the selection of optimal learning materials. Previous studies (Stoeckel et al., 2016, 2018) have demonstrated acceptable to high reliability across versions.

DOI: 10.22599/jesla.171

Each item in the NGSLT follows a specific multiple-choice format in the following three steps, and an example is shown in Figure 1.

1. Target Word: A single target word from a specific frequency band of the NGSL is presented.
2. Context Sentence: A short, non-defining sentence uses the target word naturally to help indicate its part of speech and provide context, without giving away the definition.
3. Answer Choices: Four options are provided: the correct definition/synonym and three distractors.

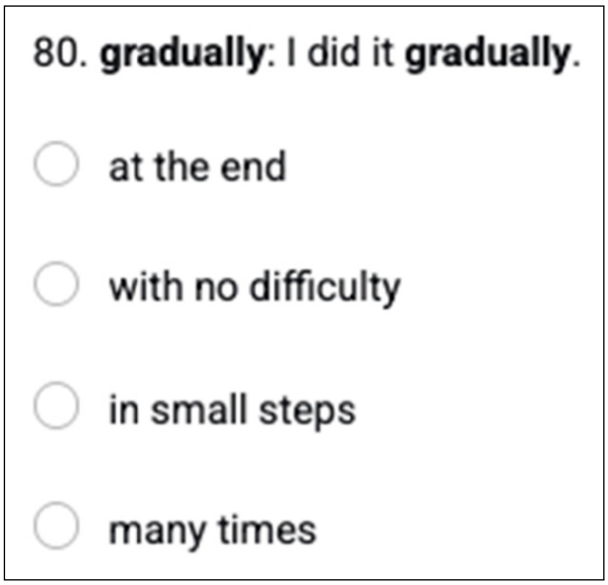



**Figure 1** Example question from the NGSLT (Version A).

Most students in the study were able to complete the NGSLT in 20–30 minutes. In this paper, the results from the first, second, and third data collection points are referred to as NGSLT1, NGSLT2, and NGSLT3, respectively.

### 2.2.2. PPLS engagement questionnaire

The level of engagement in this study was measured using the PPLS engagement questionnaire, which consists of four items (see Table 1). Responses were recorded on a six-point Likert scale, ranging from 6 (completely agree) to 1 (completely disagree). The questionnaire was originally developed by Fukada et al. (2017) to assess the extent to which English learners help each other, and it has been employed in multiple empirical studies examining PPLS engagement (e.g., Fukuda et al., 2021, 2026). These previous studies did not report reliability for the questionnaire. Thus, the reliability in the current study will be reported in the Results section.

**Table 1** Items for the PPLS engagement questionnaire.

| NUMBER | ITEM |
|---|---|
| 1 | My classmates are helping me to learn English. |
| 2 | I am helping my classmates to learn English. |
| 3 | My classmates are helping me to enjoy English. |
| 4 | I am helping my classmates to enjoy English. |

## 2.3. METHODS OF ANALYSIS

The data were analyzed using descriptive statistics, internal consistency, ANOVA, and SEM. Internal consistency was evaluated using Cronbach's alpha, with .80 adopted as the criterion for good consistency. Effect sizes for the repeated-measures ANOVA were reported using partial eta-squared ($\eta^2p$). Larger $\eta^2p$ values indicate a stronger effect of the factor on the outcome, with values of approximately .01, .06, and .14 conventionally interpreted as small, medium, and large effects, respectively (Cohen, 1988), although these thresholds should be considered relative to the research context and field-specific norms.

DOI: 10.22599/jesla.171

Among the various types of SEM, GCM was used. In GCM, an individual's initial status on the outcome variable at a specified time point is represented by the intercept, whereas the rate and direction of change in the outcome over time are represented by the slope. The model was specified as a latent-basis growth model: intercept loadings were fixed to 1.00 for all three NGSLT scores; slope loadings were fixed to 0.00 for NGSLT1 and 1.00 for NGSLT3; and the NGSLT2 slope loading was freely estimated to allow for nonlinear change. Because the first slope loading was fixed to 0.00, the intercept represents the expected baseline NGSLT score at NGSLT1, even though the intercept factor loads on all three repeated measures. Model fit was evaluated using chi-square divided by the degrees of freedom ($\chi^2$/df), the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), the comparative fit index (CFI), and the non-normed fit index (NNFI). No information-theoretic indices were reported because such indices are typically employed when alternative models are compared, which was not the focus of the current study. Model fit was considered excellent when $\chi^2$/df was below 2 (Tabachnick & Fidell, 2014), RMSEA and SRMR were below .05 (Byrne, 2016), and CFI and NNFI were above .95 (Byrne, 2016), although these values were treated as guidelines rather than absolute cutoffs.

The GCM was estimated in EQS 6.4 (Bentler & Wu, 2018) using maximum likelihood (ML) estimation on raw data. Out of the 164 participants included in the descriptive analyses, 149 had complete data on all variables included in the GCM and were retained for the analysis ($N$ = 149). The remaining 15 cases (9.1%) were excluded through listwise deletion, resulting in a complete-case dataset with no missing values. Descriptive statistics were obtained with JASP version 0.95.3 (JASP Team, 2025), whereas other traditional statistical analyses, such as internal consistency and ANOVA, were conducted with IBM SPSS Statistics (IBM Corp, 2024).

## 3. RESULTS

The results section is organized to move from descriptive to model-based evidence of vocabulary development over time. It begins with descriptive statistics and reliability information for the PPLS engagement measure and NGSLT scores, followed by analyses of mean changes across the three testing occasions. Building on these patterns, GCM is then used to capture the form of change in vocabulary development and to evaluate the role of engagement and initial status in predicting individual differences in growth.

### 3.1. DESCRIPTIVE STATISTICS

Table 2 shows the descriptive statistics of the PPLS engagement questionnaire and the three NGSLT scores for the 164 students retained for descriptive analyses. The reliability measured by Cronbach's alpha was high ($\alpha$ = .86), and none of the four items negatively affected internal consistency. Therefore, the questionnaire was considered reliable for use in this study.

| | *M* | *SD* | MIN. | MAX. |
|---|---|---|---|---|
| PPLS engagement | 4.95 | 0.79 | 1 | 6 |
| NGSLT1 | 76.68 | 11.09 | 16 | 99 |
| NGSLT2 | 87.34 | 8.95 | 59 | 100 |
| NGSLT3 | 84.48 | 10.94 | 45 | 100 |

**Table 2** Descriptive statistics for PPLS engagement and NGSLT scores (descriptive-analysis sample, $N$ = 164).

*Notes:* $M$ = mean; $SD$ = standard deviation.

The NGSLT means suggest a non-monotonic trend in vocabulary scores. They increased significantly from 76.68 for NGSLT1 in April to 87.34 for NGSLT2 in July. In December, the mean declined slightly to 84.48 for NGSLT3. This pattern motivated the latent-basis specification, as a strictly linear slope can be a poor representation when change accelerates and then levels off or reverses across occasions.

A repeated-measures ANOVA was conducted to examine changes in NGSLT scores across the three time points (NGSLT1, NGSLT2, NGSLT3), using all cases with complete data ($N$ = 164). The effect of time was statistically significant, $F(1.89, 307.46) = 94.43$, $p < .001$, with a large effect size, partial $\eta^2 = .37$. Polynomial contrasts revealed significant linear, $F(1, 163) = 85.62$, $p < .001$,

and quadratic trends, $F(1, 163) = 105.23$, $p < .001$. Estimated marginal means showed that NGSLT scores increased from Time 1 ($M = 76.68$, $SE = 0.87$) to Time 2 ($M = 87.34$, $SE = 0.70$), followed by a slight decrease at Time 3 ($M = 84.48$, $SE = 0.85$).

## 3.2. MODEL SPECIFICATION AND ESTIMATION

Vocabulary development was analyzed using a latent growth curve model estimated in EQS using the complete-case GCM sample ($N = 149$). PPLS engagement (Engagement) was included as a time-invariant predictor of the growth process. Vocabulary was measured with the NGSLT on three occasions: NGSLT1, NGSLT2, and NGSLT3. The overall model structure and standardized estimates are displayed in Figure 2.

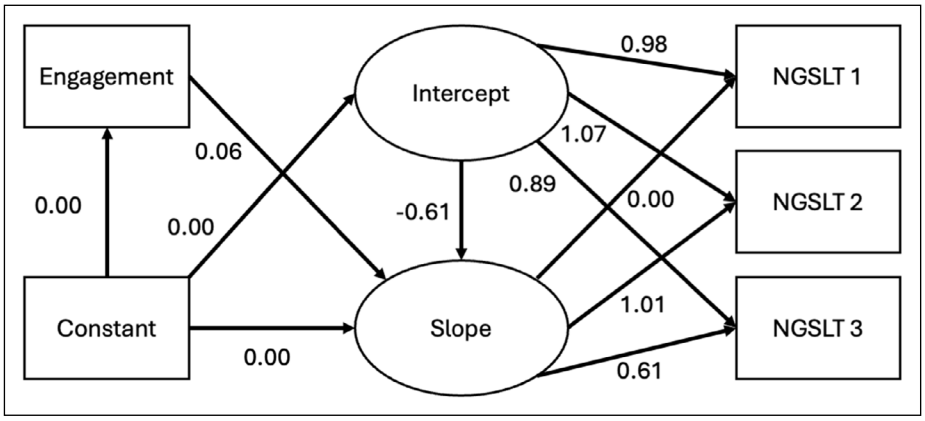



**Figure 2** Latent growth curve model with Engagement predicting vocabulary growth (standardized estimates).

*Note:* Engagement predicts the slope weakly (β = .06), while the intercept-to-slope path is notably negative (β = –.61). NGSLT2 is positioned with a freed time score (loading > 1), consistent with the observed rise from NGSLT1 to NGSLT2 followed by a decline by NGSLT3.

The growth model specified (a) a latent intercept factor (Intercept) with factor loadings fixed to 1.00 for all three NGSLT scores and (b) a latent slope factor (Slope) with the NGSLT1 loading fixed to 0.00 and the NGSLT3 loading fixed to 1.00. Although the intercept factor loads on all three repeated measures, it represents the expected baseline score because the NGSLT1 slope loading was fixed to 0.00; NGSLT2 and NGSLT3 contribute to estimating the latent baseline component and individual differences around it, but they are not treated as separate measures of initial knowledge. The NGSLT2 slope loading was freely estimated (latent-basis specification) to allow for nonlinearity in change across the three measurement occasions. With only three measurement occasions, this parameterization was deliberately parsimonious: the first and third slope loadings anchored the time metric, and only the intermediate loading was freed. Thus, the model estimates a single nonlinear trajectory rather than comparing multiple alternative growth forms. Standardized estimates for the final model are shown in Figure 2. Multivariate kurtosis was modest (Mardia normalized estimate = 1.33), suggesting that ML estimation was reasonably tenable under the normal-theory assumptions used for standard errors.

## 3.3. MODEL FIT

Global fit indices indicated excellent fit, $\chi^2(2) = 0.08$, $p = .961$, CFI = 1.00, SRMR = 0.008, RMSEA = 0.00, 90% CI [0.00, 0.14] (Table 3). At the same time, the model had only two degrees of freedom and was therefore close to saturated. Under such conditions, incremental fit indices may exceed 1.00 (e.g., CFI, NNFI) and RMSEA can be unstable. Accordingly, the fit statistics are reported for completeness, but interpretation emphasizes substantive plausibility and parameter estimates rather than treating "perfect" global fit as strongly diagnostic.

## 3.4. PARAMETER ESTIMATES

### 3.4.1. Shape of change (latent-basis time score)

The freely estimated NGSLT2 loading on the slope factor was 1.38 ($SE = 0.12$), $z = 11.34$, $p < .001$ (Table 4). Because the NGSLT3 loading was fixed at 1.00, an NGSLT2 loading greater than 1.00 implies that NGSLT2 represented a comparatively pronounced increase on the modeled growth dimension, consistent with the observed rise from NGSLT1 to NGSLT2 followed by a modest decline by NGSLT3 (Table 2).

| INDEX | VALUE | INTERPRETIVE GUIDELINE |
|---|---|---|
| $\chi^2$(df) | 0.08 (2); $\chi^2$/df = 0.04 | $\chi^2$/df < 2; lower values indicate closer fit |
| *p* | .961 | > .05 suggests no statistically significant model-data misfit |
| CFI | 1.00 | ≥ .95 is conventionally considered excellent |
| NNFI | 1.05 | ≥ .95 is conventionally considered excellent; values can exceed 1.00 with low df |
| SRMR | 0.01 | ≤ .05 is conventionally considered excellent |
| RMSEA | 0.00 | ≤ .05 is conventionally considered excellent |
| 90% CI for RMSEA | [0.00, 0.14] | Lower/narrower intervals are preferable; unstable with df = 2 |

**Table 3** Model fit indices (ML estimation).

*Note:* The interpretive guidelines are conventional heuristics, not absolute rules. With df = 2, incremental fit indices can exceed 1.0 and RMSEA can be unstable; therefore, these indices are reported for transparency and interpreted cautiously rather than as definitive evidence of model adequacy.

| PARAMETER | UNSTANDARDIZED (b) | SE | z | STANDARDIZED |
|---|---|---|---|---|
| NGSLT2 slope loading | 1.38 | 0.12 | 11.34 | 1.01 |
| Slope ← Engagement | 0.49 | 0.61 | 0.81 | 0.06 |
| Slope ← Intercept | −0.42 | 0.13 | −3.32 | −0.61 |
| Mean (Intercept) | 76.73 | 0.81 | 94.54 | – |
| Mean (Slope) | 7.62 | – | – | – |
| Var (Intercept) | 93.88 | 29.36 | 3.20 | – |
| Residual Var (Slope) | 27.48 | 8.11 | 3.39 | – |
| $R^2$ (Slope) | | | | 0.38 |

**Table 4** Selected parameter estimates from the latent growth curve model.

*Note:* Mean (Intercept) is the model-implied baseline NGSLT score. Mean (Slope) is the model-implied average change from NGSLT1 to NGSLT3 because the slope loadings were fixed to 0.00 and 1.00 at those occasions; it should not be confused with the uncentered conditional constant in the EQS slope equation. Residual Var (Slope) is the unexplained variance in the slope factor after Engagement and Intercept are included. The standardized column reflects EQS standardized solution values where available. Dashes indicate values that were not applicable or were not directly reported in the standardized solution.

### 3.4.2. Predictors of growth (slope factor)

Two predictors, Engagement and Intercept, were specified for Slope. Engagement did not show a statistically reliable association with Slope, $b$ = 0.49, $SE$ = 0.61, 95% CI [−0.69, 1.68], $z$ = 0.81, $p$ = .416, β = .06. In contrast, the Intercept-to-Slope path was negative and statistically significant, $b$ = −0.42, $SE$ = 0.13, $z$ = −3.32, $p$ < .001, β = −.61, indicating that students with higher vocabulary scores at the first data collection point tended to exhibit smaller subsequent gains (or comparatively flatter change) across the year.

### 3.4.3. Variances (individual differences)

The mean of the slope factor was positive ($M$ = 7.62), indicating that the average NGSLT3 score remained above the NGSLT1 baseline in the EQS analytic sample. Significant variance was also observed for the Intercept factor and for the residual variance of the Slope factor, indicating meaningful between-student differences in vocabulary size at the first point of testing and remaining individual differences in change over time after accounting for Engagement and Intercept: Var (Intercept) = 93.88, $SE$ = 29.36, $z$ = 3.20; Residual Var (Slope) = 27.48, $SE$ = 8.11, $z$ = 3.39 (Table 4). The predictors accounted for 37.6% of the variance in the slope factor ($R^2$ = .38).

Overall, the results provide converging descriptive and model-based evidence that vocabulary development over the academic year was characterized by substantial individual differences and a clearly nonlinear pattern of change. While mean NGSLT scores increased markedly early in the year, subsequent growth leveled off, a pattern well captured by the GCM. Engagement did not emerge as a significant predictor of vocabulary growth, whereas initial vocabulary status was strongly and negatively related to subsequent change, suggesting a compensatory pattern in which students with lower initial scores tended to show greater gains.

## 4. DISCUSSION

The present study examined longitudinal changes in university students' vocabulary knowledge using the NGSLT (Stoeckel & Bennett, 2015) and tested whether PPLS engagement predicted this change within a growth curve model. It addressed a gap established in the literature review: only a very limited number of studies have examined engagement as a predictor of vocabulary development (e.g., Oikawa & Uchihara, 2026; Zhang et al., 2023), and no previous

DOI: 10.22599/jesla.171

study has examined PPLS engagement in relation to vocabulary growth. The results complicate the expected picture in two ways: growth was heavily front-loaded and uneven across learners, and the anticipated engagement–growth association did not appear.

Regarding RQ1, NGSLT scores did not rise steadily. The mean increased from 76.68 in April to 87.34 in July, then fell to 84.48 in December, and the freely estimated loading for the second occasion (1.38) indicates that on the modeled growth dimension, students had gained more by July than they retained by the end of the year. Much vocabulary research has addressed the construction of word lists and the coverage they provide (Coxhead, 2000; Webb & Rodgers, 2009) and the measurement of learners' vocabulary size (Nation & Beglar, 2007); studies that model L2 vocabulary growth across three or more measurement occasions remain scarce (e.g., De Wilde, 2025). The present data add a specific observation to that small body of work: even for the highest-frequency vocabulary of English, essentially all of the year's measurable net gain occurred in the first semester, with the mean trajectory overshooting its December endpoint by more than a third.

Two explanations for this shape need to be weighed against each other, and the design does not fully separate them. The first is curricular. Students spent nine hours per week in English courses and were required to study a selected list of 400 NGSL words, with weekly quizzes. To the extent that this list was front-loaded, the April–July surge partly reflects direct instruction on the same frequency bands that the NGSLT samples. The second explanation is test form. All students took Form A in April, Form C in July, and Form B in December, so measurement occasion and form are confounded. The December decline is interpretable as attrition of knowledge only if Forms B and C are of equal difficulty, and, to our knowledge, the three English forms have not been formally equated. Nadasdy et al. (2018) found that reliability and criterion correlations differed across versions, which cautions against treating the forms as interchangeable. Counterbalancing form order or linking the forms through common items would allow future administrations to separate real within-year change from form effects.

For RQ2, the model confirmed large individual differences on both growth factors, and their size is worth stating in test units. The intercept variance of 93.88 corresponds to a standard deviation of approximately 10 points on the 100-point test, indicating considerable variation in students' initial vocabulary knowledge. The residual slope variance of 27.48 ($SD$ = 5.24) means that even after initial level and engagement were accounted for, students' April-to-December change still differed by several points in either direction. At the test's sampling ratio of roughly 28 headwords per item, a five-point difference in yearly growth corresponds to approximately 140 NGSL words. Comparable heterogeneity has been reported for affective variables modeled with growth curves (Elahi Shirvan & Taherian, 2021; Li & Zhang, 2023); the present study documents it for a lexical outcome, where individual differences translate directly into words known.

With respect to RQ3, the results indicated that PPLS engagement did not significantly predict vocabulary growth in the current model. The contrast with the engagement–achievement literature is worth quantifying. Okunuki and Kashimura's (2024) meta-analysis reports an overall correlation of $r$ = .23 between engagement and achievement in L2 learning, with $r$ = .24 for behavioral engagement specifically. The present standardized estimate was .06, with a 95% confidence interval around the unstandardized coefficient of –0.69 to 1.68 ($b$ = 0.49, $SE$ = 0.61). Even the largest direct effect compatible with that interval, just under 1.7 NGSLT points of yearly growth per scale point of engagement (about 1.3 points per standard deviation of the measure), would be small against a mean gain of 7.62 points. The data do not merely fail to confirm a direct effect; they constrain any direct effect to be modest. How engagement was operationalized matters for interpreting this result. PPLS engagement was measured with four items asking whether students and their classmates were helping one another learn and enjoy English (Fukada et al., 2017), a social and behavioral form of engagement distinct from the cognitive dimension. Zhang et al. (2023), by contrast, measured engagement during digital game-based vocabulary learning across all three factors of Fredricks et al.'s (2004) model: behavioral, emotional, and cognitive. In their study, overall engagement (the three factors combined) predicted learning and retention, but behavioral engagement alone showed mixed effects, with heavy investment of time sometimes accompanying reduced efficiency. Read against that finding, a weak direct path from a behavioral, peer-oriented measure to vocabulary scores is less anomalous than it first appears.

DOI: 10.22599/jesla.171

Three features of the present design offer specific explanations for this null result. First, timing: engagement was measured once, in June, and treated as time-invariant, whereas the outcome spans April to December. The June measure may therefore reflect peer relationships that had already developed during the first part of the academic year, and it cannot capture either students' initial levels of PPLS engagement or subsequent changes in those relationships. Fukada et al. (in press) found that the association between PPLS engagement and motivation strengthened over a semester, and a single mid-year snapshot cannot register such reciprocal development. Second, restriction of range: the sample mean was 4.95 on a six-point scale (*SD* = 0.79), so most students endorsed the items, and the predictor had little variance to share with the growth factor. Third, distance between construct and outcome: in this program, NGSL vocabulary was learned chiefly through individual study of bilingual word lists, Quizlet practice, and weekly quizzes. Peer support plausibly shapes outcomes that depend on interaction, such as motivation, persistence, and willingness to use the language, before it shapes performance on an individually studied word list. None of these considerations supports the existence of a direct effect. The straightforward reading is that a direct effect on this particular outcome is either absent or too small to detect here, while indirect pathways remain untested.

If PPLS engagement affects vocabulary at all, the route suggested by the socio-dynamic framing in which the construct originates (Murphey et al., 2014; Vygotsky, 1978) runs through motivation: Ideal Classmates Priming raises PPLS engagement, engagement feeds temporal self-image motivation (Fukada et al., in press), and motivation sustains the self-regulated study through which vocabulary from the target list is actually acquired. This chain is testable. It requires engagement and candidate mediators, such as study time, weekly quiz performance, and vocabulary-specific self-efficacy (Furusho & Kitagami, 2022), to be measured at each collection point alongside the NGSLT and modeled as parallel processes rather than as a single direct effect. Affective variables such as ambiguity tolerance and reading anxiety (Baharloo & Ghasemian, 2024) become relevant where learning is reading-based.

Regarding RQ4, initial status was the strongest predictor in the model: each additional point on the first measurement occasion predicted 0.42 fewer points of the first-to-third point gain ($\beta$ = -.61). Before this is read as compensatory learning, two artifacts deserve attention. The first is the test ceiling: the July mean of 87.34 already sits above the 80–85% band that the NGSLT documentation treats as mastery, and students who began near 90 had almost no room to gain. The second is regression to the mean: with imperfectly reliable measures, unusually high April scores partly reflect positive measurement error that will not recur at later occasions (Field, 2018; Kline, 2023), and with unequated forms this cannot be separated from genuine convergence. What survives these caveats is still informative. Lower-scoring students gained most on exactly the bands the curriculum targeted (the 400-word list was built from NGSL words students were unlikely to know), suggesting the first-year vocabulary strand works as intended for the students who need it, while adding little measurable knowledge for those who arrive near mastery. For that latter group, continued NGSL-focused assessment yields diminishing information; a measure targeting mid-frequency and academic vocabulary, for instance one built from the Global Academic Vocabulary list already used in the program (Wadden et al., 2016), would be more sensitive to their growth.

Methodologically, the study shows what growth modeling added over the occasion-by-occasion comparison, and where its limits lay in this application. The model separated where students started from how much they changed, quantified individual differences in both, and captured the shape of change through the freely estimated second loading, none of which the repeated-measures ANOVA could do. GCM remains rare in vocabulary research despite explicit calls for its wider use in applied linguistics (Zhou et al., 2023) and accessible methodological guidance (Curran et al., 2010; Parsons & McCormick, 2024), and this application illustrates both its payoff and its data demands. With three occasions the model has two degrees of freedom and is nearly saturated, so the excellent global fit is expected rather than diagnostic; the parameter estimates, not the fit indices, carry the evidential weight. Their robustness should be examined with additional data collection points, alternative codings of time, and, most consequentially for this instrument, equating of the test forms.

The findings describe a first year in which vocabulary growth was front-loaded, strongly conditioned on starting level, and in which PPLS engagement, as measured here, did not directly predict that growth. This does not diminish the pedagogical case for engagement; it argues

DOI: 10.22599/jesla.171

against expecting a single mid-year measure of peer support to translate straightforwardly into receptive vocabulary scores. Whether it translates indirectly is a design question, and the design that would answer it is within reach of an ordinary first-year program: repeated engagement measurement, explicit mediators, and equated test forms.

## 5. CONCLUSION

This study investigated longitudinal vocabulary development and its relationship with learners' PPLS engagement among first-year Japanese university students by integrating NGSLT-based vocabulary measurement with GCM. By tracking vocabulary knowledge at three points across an academic year and modeling individual growth trajectories, the study demonstrated that vocabulary development was nonlinear and characterized by substantial individual differences. Vocabulary gains were most pronounced in the first half of the academic year, followed by a modest decline, a pattern well captured by a latent-basis growth model. Notably, vocabulary scores at the first data collection point emerged as a strong predictor of subsequent change, with lower-proficiency students showing comparatively greater gains, highlighting the importance of considering learners' starting points when examining vocabulary growth.

Contrary to expectations and to some findings in the broader engagement–achievement literature, PPLS engagement did not directly predict vocabulary growth in the present model. One possible explanation is that PPLS engagement represents a conceptually distinct form of engagement from that examined in much previous research. The PPLS scale focuses specifically on students' perceptions of mutually supportive peer learning—whether students help and are helped by classmates to learn and enjoy English—rather than on broader cognitive, behavioral, and emotional dimensions of engagement. Therefore, the lack of a direct relationship with vocabulary growth should not necessarily be interpreted as inconsistent with previous engagement research. Instead, the influence of PPLS engagement on vocabulary development may be indirect, context-dependent, or mediated through more proximal learning behaviors.

Three limitations should be acknowledged. First, as noted earlier, PPLS engagement was measured only once and treated as a time-invariant predictor, which restricts the ability to capture its potentially dynamic nature. Engagement is likely to fluctuate across instructional phases and in response to classroom experiences; therefore, collecting PPLS engagement data at multiple time points would allow future studies to model its development in parallel with vocabulary growth and to examine time-varying effects. Second, the study relied on a single measure of engagement focused on prosocial peer learning support, whereas engagement is commonly conceptualized as a multidimensional construct comprising three components. Incorporating this three-factor model into the current framework may provide a more comprehensive understanding of the relationship between engagement and vocabulary development and may reveal associations that were not captured by the present measure. Third, with only three measurement occasions and a model close to saturation, the robustness of the growth parameters would benefit from replication with additional data collection points and alternative model specifications.

Despite these limitations, the present study contributes to SLA research by being among the first to integrate vocabulary development and engagement within a single growth modeling framework. Methodologically, it illustrates the value of GCM for capturing nonlinear change and individual variability in vocabulary learning. Substantively, it highlights that vocabulary growth is not uniform across learners and that initial proficiency plays a critical role in shaping developmental trajectories. Future research that combines richer, time-sensitive measures of engagement with longitudinal vocabulary assessment will be essential for clarifying how social and psychological factors interact with language development over time.

## DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE MANUSCRIPT PREPARATION PROCESS

After the manuscript had been written by the authors, AI-based writing assistance tools, specifically ChatGPT (OpenAI, 2026) and Grammarly (Grammarly, 2026), were used solely to check grammar and spelling. The content, analysis, and interpretations presented in this paper are the authors'.

## AUTHOR AFFILIATIONS

**Tetsuya Fukuda** orcid.org/0000-0003-2117-1725
Juntendo University, JP
**Paul Wadden** orcid.org/0009-0005-0546-321X
Juntendo University, JP
**Siwon Park** orcid.org/0009-0004-5561-4162
Juntendo University, JP
**Marcel Van Amelsvoort** orcid.org/0009-0000-6177-0701
Juntendo University, JP

## REFERENCES

**Alexander, K. L., Entwisle, D. R.,** & **Horsey, C. S.** (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*(2), 87–107. https://doi.org/10.2307/2673158

**Baharloo, A.,** & **Ghasemian, Z.** (2024). The interplay of Iranian EFL learners' tolerance of ambiguity and their reading performance encountering unknown vocabularies: Taking their L2 reading anxiety into account. *International Journal of Research in English Education*, *9*(3), 61–74.

**Bentler, P. M.,** & **Wu, E. J. C.** (2018). *EQS for Windows* (Version 6.4) [Computer software]. Multivariate Software, Inc. https://mvsoft.com/eqsdownload

**Browne, C.** (2014). A new general service list: The better mousetrap we've been looking for. *Vocabulary Learning and Instruction, 3*(2), 1–10. https://doi.org/10.7820/vli.v03.2.browne

**Byrne, B. M.** (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge. https://doi.org/10.4324/9781315757421

**Cohen, J.** (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

**Coxhead, A.** (2000). A new academic word list. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

**Csizér, K.,** & **Dörnyei, Z.** (2005). Language learners' motivational profiles and their motivated learning behavior. *Language Learning, 55*(4), 613–659. https://doi.org/10.1111/j.0023-8333.2005.00319.x

**Curran, P. J., Obeidat, K.,** & **Losardo, D.** (2010). Twelve frequently asked questions about growth curve modeling. *Journal of Cognition and Development, 11*(2), 121–136. https://doi.org/10.1080/15248371003699969

**De Wilde, V.** (2025). Adolescent learners' L2 English vocabulary knowledge and contact with extramural English: Longitudinal development and relationships between L2 vocabulary and extramural English. *Bilingualism: Language and Cognition*. Advance online publication. https://doi.org/10.1017/S1366728925100540

**Elahi Shirvan, M.,** & **Taherian, T.** (2021). Longitudinal examination of university students' foreign language enjoyment and foreign language classroom anxiety in the course of general English: Latent growth curve modeling. *International Journal of Bilingual Education and Bilingualism, 24*(1), 31–49. https://doi.org/10.1080/13670050.2018.1441804

**Elahi Shirvan, M., Taherian, T., Shahnama, M.,** & **Yazdanmehr, E.** (2021). A longitudinal study of foreign language enjoyment and L2 grit: A latent growth curve modeling. *Frontiers in Psychology, 12*, 720326. https://doi.org/10.3389/fpsyg.2021.720326

**ETS.** (2025). *TOEFL ITP: Test and score data summary*.

**Field, A.** (2018). *Discovering statistics using IBM SPSS statistics*. Sage.

**Fredricks, J. A., Blumenfeld, P. C.,** & **Paris, A. H.** (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

**Fukada, Y., Falout, J.,** & **Fukuda, T.** (in press). Ideal Classmates Priming: How does Prosocial Engagement increase temporal self-image motivations? In A. Leis, N. Zarrinabadi, & Y. Yao (Eds.), *Intervention studies in motivating the language learner*. Palgrave Macmillan.

**Fukada, Y., Fukuda, T., Falout, J.,** & **Murphey, T.** (2017). Collaboratively visualizing possible others. *Learner Development Journal, 1*(1), 79–93.

**Fukuda, T.** (2020). *School belonging and L2 motivation of first-year students at four Japanese universities* (Publication No. 28027020) [Doctoral dissertation, Temple University]. ProQuest Dissertations and Theses Global.

**Fukuda, T., Fukada, Y., Falout, J.,** & **Murphey, T.** (2021). How ideal classmates priming increases EFL prosocial engagement. In P. Hiver, A. H. Al-Hoorie, & S. Mercer (Eds.), *Student engagement in the language classroom* (pp. 182–201). Multilingual Matters. https://doi.org/10.2307/jj.22730722.15

**Fukuda, T., Fukada, Y., Munezane, Y., Falout, J.,** & **Murphey, T.** (2026). Prosocial engagement, individual-level motivations, and group dynamics: How does ideal classmates priming work now? *Journal of Psychology of Language Learning, 8*(1), 205–242. https://doi.org/10.52598/jpllsi/8/1/7

**Fukuda, T., Park, S., Wadden, P., Van Amelsvoort, M.,** & **Umehara, H.** (2025). Investigation into the reliability of the New General Service List Test (NGSLT) and the relationship between NGSLT and TOEFL among Japanese university students. *Juntendo Journal of Global Studies, 10*, 43–50.

**Furusho, T.,** & **Kitagami, S.** (2022). The effects of foreign language learners' motivation on efficacy and anxiety in vocabulary learning. *Proceedings of the Japanese Society for Cognitive Psychology, 20*, 62–62. https://doi.org/10.14875/cogpsy.2022.0_62

**Gu, L.,** & **Sun, P.** (2021). Book review: Student engagement in the language classroom. *Frontiers in Psychology, 12*, 713998. https://doi.org/10.3389/fpsyg.2021.713998

**IBM Corp.** (2024). *IBM SPSS Statistics for Macintosh* (Version 29.0) [Computer software]. IBM Corp.

**JASP Team.** (2025). *JASP* (Version 0.95.3) [Computer software].

**Kline, R. B.** (2023). *Principles and practice of structural equation modeling* (5th ed.). The Guilford Press.

**Li, C.,** & **Zhang, L. J.** (2023). The development of accuracy and fluency in second language (L2) speaking related to self-efficacy through online scaffolding: A latent growth curve modeling analysis. *Journal of Psycholinguistic Research, 52*(5), 1371–1395. https://doi.org/10.1007/s10936-023-09950-7

**Markus, H.,** & **Nurius, P.** (1986). Possible selves. *American Psychologist, 41*(9), 954–969. https://doi.org/10.1037/0003-066X.41.9.954

**Murphey, T.,** & **Arao, H.** (2001). Reported belief changes through near peer role modeling. *Tesl-Ej, 5*(3), 1–15.

**Murphey, T., Falout, J., Fukuda, T.,** & **Fukada, Y.** (2014). Socio-dynamic motivating through idealizing classmates. *System, 45*, 242–253. https://doi.org/10.1016/j.system.2014.06.004

**Nadasdy, P., Aizawa, K.,** & **Iso, T.** (2018). Testing the reliability of the New General Service List Test (NGSLT) in order to better evaluate Japanese university students' written receptive vocabulary levels. *Future-Proof CALL: Language Learning as Exploration and Encounters*, 220–224. https://doi.org/10.14705/rpnet.2018.26.840

**Nation, I. S. P.,** & **Beglar, D.** (2007). A vocabulary size test. *The Language Teacher, 31*(7), 9–13. https://doi.org/10.26686/wgtn.12552197.v1

**Nation, I. S.,** & **Webb, S. A.** (2011). *Researching and analyzing vocabulary*. Heinle Cengage Learning.

**Oikawa, G.,** & **Uchihara, T.** (2026). Learner engagement in L2 vocabulary learning through communicative tasks: Comparing learner- and teacher-generated content. *TESOL Quarterly*. Advance online publication. https://doi.org/10.1002/tesq.70166

**Okunuki, A.,** & **Kashimura, Y.** (2024). Student engagement and academic achievement in second language learning: A meta-analysis. *JACET Journal, 68*, 71–90.

**Pan, J., Liang, J.,** & **Hu, Y.** (2026). Gamified behavioral engagement as an indirect association linking ideal L2 self and vocabulary achievement in mobile-assisted EFL vocabulary learning. *Frontiers in Psychology, 17*, 1836665. https://doi.org/10.3389/fpsyg.2026.1836665

**Park, S., Fukuda, T.,** & **Wadden, P.** (2025). The mediating role of vocabulary in linking L1 reading and L2 proficiency. *International Journal of TESOL Studies, 251107*, 1–18. https://doi.org/10.58304/ijts.251107

**Parsons, S.,** & **McCormick, E. M.** (2024). Limitations of two time point data for understanding individual differences in longitudinal modeling—what can difference reveal about change? *Developmental Cognitive Neuroscience, 66*, 101353. https://doi.org/10.1016/j.dcn.2024.101353

**Plonsky, L.** (2014). Study quality in quantitative L2 research (1990–2010): A methodological synthesis and call for reform. *The Modern Language Journal, 98*(1), 450–470. https://doi.org/10.1111/j.1540-4781.2014.12058.x

**Purpura, J. E.** (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge University Press.

**Reeve, J.,** & **Tseng, C.** (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology, 36*(4), 257–267. https://doi.org/10.1016/j.cedpsych.2011.05.002

**Stoeckel, T.,** & **Bennett, P.** (2015). A test of the new General Service List. *Vocabulary Learning and Instruction, 4*(1), 1–8. https://doi.org/10.7820/vli.v04.1.stoeckel.bennett

**Stoeckel, T., Bennett, P.,** & **Ishii, T.** (2016). *New General Service List test: Updated documentation*. [Technical report]. https://www.academia.edu/30266171/New_General_Service_List_Test_Updated_Documentation

**Stoeckel, T., Ishii, T.,** & **Bennett, P.** (2018). A Japanese-English bilingual version of the new general service list test. *JALT Journal, 40*(1), 5–21. https://doi.org/10.37546/JALTJJ40.1-1

**Tabachnick, B. G.,** & **Fidell, L. S.** (2014). *Using multivariate statistics*. Pearson.

**Vygotsky, L. S.** (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

**Wadden, P., Ferreira, D.,** & **Rush, E.** (2016). The global academic vocabulary lexicon: A new ELT resource. *Accents Asia, 8*(2), 135–142.

**Wang, M., Willett, J. B.,** & **Eccles, J. S.** (2011). The assessment of school engagement: Examining dimensionality and measurement invariance by gender and race/ethnicity. *Journal of School Psychology*, *49*(4), 465–480. https://doi.org/10.1016/j.jsp.2011.04.001

**Webb, S.,** & **Rodgers, M. P.** (2009). Vocabulary demands of television programs. *Language Learning, 59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

**West, M.** (1953). *A general service list of English words*. Longman.

**Willms, J. D.** (2003). *Student engagement at school: A sense of belonging and participation*. OECD.

**Winke, P.** (2014). Testing hypotheses about language learning using structural equation modeling. *Annual Review of Applied Linguistics, 34*, 102–122. https://doi.org/10.1017/S0267190514000075

**Zhang, R., Zou, D.,** & **Cheng, G.** (2023). Learner engagement in digital game-based vocabulary learning and its effects on EFL vocabulary development. *System, 119*, 103173. https://doi.org/10.1016/j.system.2023.103173

**Zhou, S., Hiver, P.,** & **Al-Hoorie, A. H.** (2021). Measuring L2 engagement: A review of issues and applications. In P. Hiver, A. H. Al-Hoorie, & S. Mercer (Eds.), *Student engagement in the language classroom* (pp. 75–98). Multilingual Matters. https://doi.org/10.2307/jj.22730722.10

**Zhou, S. A., Yang, Y.,** & **Hiver, P.** (2023). Individual growth curve modeling for applied linguistics research. *Research Methods in Applied Linguistics, 2*(1), 100043. https://doi.org/10.1016/j.rmal.2023.100043

**TO CITE THIS ARTICLE:**
Fukuda, T., Wadden, P., Park, S., & Van Amelsvoort, M. (2026). An investigation of vocabulary development and prosocial peer learning support engagement using Growth Curve Modeling. *Journal of the European Second Language Association*, 10(1), 165–180. DOI: https://doi.org/10.22599/jesla.171

**Submitted:** 01 February 2026
**Accepted:** 01 October 2026
**Published:** 12 November 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.