



Narrow Reading with AI-Generated Texts for L2 Vocabulary Learning: Exploring the Roles of Vocabulary Size and Working Memory*

Eun Young Kang (Kongju National University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: July 9, 2026

Revised: July 29, 2026

Accepted: July 30, 2026

Kang, Eun Young
Professor, Division of Liberal Arts
Kongju National University
1223-24, Cheonan-daero, Seobuk-
gu, Cheonan-si, Chungcheongnam-
do, Korea
Tel: +82-41-850-8660
Email: ekang@kongju.ac.kr

* This work was supported by the research grant of Kongju National University in 2025.

ABSTRACT

Kang, Eun Young. 2026. Narrow reading with AI-generated texts for L2 vocabulary learning: Exploring the roles of vocabulary size and working memory. *Korean Journal of English Language and Linguistics* 26, 1164-1181.

This study examined theme-based narrow reading (NR) with AI-generated texts for L2 vocabulary learning and whether vocabulary size and working memory (WM) predicted gains. In a quasi-experimental pretest-posttest design, 96 Korean EFL university students were assigned to a narrow reading or comparison group. Over six sessions, the NR group read AI-generated texts on one topic (measles outbreak), while the comparison group read texts on unrelated topics. Knowledge of eight target words was assessed through meaning recall and form recognition tests at pretest, immediate posttest, and delayed posttest. A repeated measures ANCOVA revealed a significant main effect of group, with the NR group outperforming the comparison group at both time points after controlling for pretest scores. The NR group's advantage was maintained over a two-week delay. Regression analyses within the NR group showed that both vocabulary size and WM predicted immediate gains, but only WM predicted delayed gains, suggesting a more enduring role in retention. These findings demonstrate that AI-generated texts offer a practical, pedagogically principled alternative to authentic NR materials, enabling precise control over vocabulary frequency and thematic coherence. The results also underscore the importance of WM—as a driver of learners' attention to target lexis—in vocabulary learning.

KEYWORDS

narrow reading, AI-generated texts, L2 vocabulary learning, EFL, individual differences

1. Introduction

Reading is a primary source of input for second language (L2) development, and consistent exposure to meaningful texts has long been proposed as a key means of enhancing L2 learning (Ahmed and Han 2018). Among the pedagogical strategies developed to maximize such exposure, narrow reading (NR)—in which learners read multiple texts on the same topic, within the same genre, or by the same author (Krashen 2004)—has received sustained attention. Theme-based NR in particular is thought to promote vocabulary acquisition through the repeated recycling of topic-related lexis (Gardner 2008), potentially enabling learners to encounter target words across different yet contextually related texts. For instance, a learner reading multiple texts on the topic of climate change would repeatedly encounter domain-specific vocabulary such as *emissions*, *drought*, and *glacier* across different texts, reinforcing their form-meaning mappings with each encounter. This could reduce cognitive load during reading and enhance noticing of new lexical items, both of which are critical conditions for vocabulary learning.

Despite its pedagogical promise, theme-based NR faces a persistent practical challenge: sourcing texts that contain abundant thematically core vocabulary—words closely related to the topic that recur across multiple texts. While graded readers and books organized around a single theme have been recommended as potential sources, they are often inaccessible due to cost or copyright restrictions. Online news dealing with a single topic have been proposed as a more accessible alternative (Dang and Lu 2025), yet authentic materials are difficult to calibrate for target vocabulary frequency or learner proficiency.

Artificial intelligence (AI) text generation offers a principled solution to these limitations. AI-generated texts can be produced rapidly, tailored to a specific proficiency level and thematic focus, and structured to ensure that target vocabulary appears at sufficient frequencies across texts—conditions that are difficult to guarantee with authentic sources (Koraishi 2023). Indeed, AI tools have seen growing use in language education, and recent evidence suggests that AI-adapted texts can improve reading comprehension and inferencing among EFL learners (Çelik et al. 2024). However, their application to NR material development remains unexplored. Recent systematic reviews confirm that reading has received comparatively less attention than writing as a focus of AI integration in language learning (Lee et al. 2026). The present study addresses both gaps by employing Edcafe (<https://edcafe.ai>), an AI-powered text generation platform, to produce thematically coherent reading texts as a part of a theme-based NR intervention—representing a novel methodological contribution to NR research and offering an accessible model for practitioners.

Beyond the question of material sourcing, the conditions under which NR yields optimal vocabulary learning outcomes have also received limited empirical attention. Individual differences—specifically L2 vocabulary size and working memory (WM) capacity—are well-established predictors of incidental vocabulary acquisition through reading (Webb and Chang 2015, Teng 2024, Linck et al. 2014). Yet whether these advantages extend to theme-based NR implemented with AI-generated texts is less straightforward. Such texts can include systematic repetition of target vocabulary across multiple texts, potentially providing all learners with sufficient exposure to consolidate unfamiliar items—regardless of their vocabulary size or WM capacity. Whether individual differences continue to predict vocabulary gains under these conditions remains an open empirical question. These two considerations are not independent: the systematic repetition afforded by AI-generated texts multiplies opportunities to encounter—and potentially notice—target words, but whether learners capitalize on those opportunities depends on their cognitive resources. The present study therefore treats input design and learner-internal factors as jointly shaping how learners attend to vocabulary during reading.

Theme-based NR is of particular interest from a language awareness perspective, because encountering the same target words repeatedly across thematically related texts creates recurring opportunities for learners to notice and attend to those items as salient features of the input (Schmidt 1990). The extent to which learners capitalize on these noticing opportunities, however, may depend on learner-internal resources—particularly the attentional and executive control captured by working memory. The present study therefore examines vocabulary learning through narrow reading with AI-generated texts and the extent to which L2 vocabulary size and WM capacity predict learners' gains. In doing so, it makes two contributions to the field: it demonstrates the feasibility and pedagogical value of AI-generated texts as NR materials, and it advances understanding of how theme-based NR interacts with learner-internal cognitive variables that govern attention to lexical form to facilitate vocabulary development in digital L2 learning contexts.

2. Literature Review

2.1 Theme-Based NR and L2 Vocabulary Learning

Several studies have demonstrated the vocabulary learning benefits of theme-based NR. Kang (2015) found that Korean EFL high school students who read thematically related online news articles significantly outperformed a control group on both receptive and productive vocabulary measures. Dang and Lu (2025) reported similar findings with Chinese EFL university students engaged in theme-based reading on a single topic, with the NR group achieving significantly higher vocabulary gains than the control group. These findings are consistent with theoretical accounts emphasizing that repeated, contextualized encounters with words reinforce form-meaning connections and promote deeper lexical processing (Nation 2013, Verspoor and Schmitt 2013).

However, both studies relied on small sets of authentic texts delivered in printed form, limiting ecological validity in an era of widespread digital reading. Kang (2025) addressed this by implementing theme-based NR digitally via Google News, better reflecting real-world online reading practices. Nevertheless, such NR practice with authentic materials may not be feasible for all learners, particularly those at lower proficiency levels. A more fundamental challenge also concerns the difficulty of controlling target word frequency in authentic materials, as naturally occurring texts vary considerably in their lexical profiles. Target word frequency is a well-established facilitator of incidental vocabulary learning (Webb 2020), yet ensuring sufficient and repeated exposure to specific words is difficult when using authentic sources. In practice, teachers therefore struggle to prepare level-appropriate NR materials that systematically recycle target vocabulary across texts.

AI text generation presents a promising alternative to authentic NR materials. By employing a platform such as Edcafe, teachers can produce thematically coherent digital texts with target vocabulary embedded at consistent frequencies and appropriate readability levels. It should be noted, however, that findings from NR studies using authentic materials may not generalize straightforwardly to AI-generated texts. Authentic texts vary naturally in lexical density, discourse structure, and redundancy, and part of the benefits observed in previous NR studies may derive from such natural variation; conversely, the deliberately engineered repetition in AI-generated texts may alter the distribution and contextual richness of word encounters. Whether the vocabulary learning benefits of theme-based NR are preserved when texts are generated by AI therefore remains an empirical question, one that the present study directly addresses.

2.2 The Role of Vocabulary Size and Working Memory in L2 Vocabulary Learning through Reading

Vocabulary size—the estimated number of words for which learners possess at least partial knowledge of form-meaning relationships—is a well-established predictor of both reading comprehension and incidental vocabulary acquisition through reading (Laufer and Aviad-Levitzky 2017, Miralpeix and Muñoz 2018). Webb and Chang (2015) found that EFL high school students with larger vocabularies retained significantly more new words during extensive reading of graded readers. Teng (2024) similarly reported that vocabulary size was a strong predictor of form and meaning recognition in online reading among university students. The role of vocabulary size has been further confirmed across multiple input modalities, including listening and viewing (Feng and Webb 2020).

Within theme-based NR contexts, however, the role of vocabulary size is less clear. The thematic repetition inherent in NR already provides learners with multiple encounters with topic-related vocabulary across texts. When this repetition is further supported by AI-generated texts—which can be tailored to maintain thematic coherence across multiple texts while ensuring consistent exposure to target vocabulary—even learners with smaller vocabularies may receive sufficient, structured exposure to consolidate target items. Whether vocabulary size nonetheless predicts learning outcomes in this context remains an open question.

Beyond vocabulary size, WM plays a theoretically central role in L2 vocabulary acquisition (Li 2023). Defined as a limited-capacity cognitive system responsible for the temporary storage and manipulation of information during complex tasks (Baddeley 2003), WM is thought to facilitate the encoding and integration of new lexical items and their features during reading. Crucially, WM is closely tied to the attentional processes underlying noticing: learners with greater executive control are better able to allocate attention to unfamiliar lexical forms embedded in running text and to hold them in focus long enough for deeper processing (Schmidt 1990, Robinson 2003). In this sense, WM can be understood not merely as storage capacity but as an individual-difference variable that shapes how much learners consciously attend to target vocabulary during reading. Despite the theoretically established role of WM in vocabulary learning through reading, its empirical contribution in digital reading contexts specifically remains unclear. For instance, Varol and Erçetin (2016) and Ünal (2023) both found no significant WM effects on vocabulary gains among EFL learners reading electronic texts. However, such null findings may reflect measurement limitations rather than a true absence of WM effects. Both studies relied on backward digit span tasks, which primarily assess storage capacity and may lack sensitivity to the executive functioning relevant to vocabulary acquisition (de Leeuw et al. 2014). This interpretation is further supported by Linck et al.'s (2014) meta-analysis, which confirmed that the executive control component of WM, assessed through complex span tasks, is a stronger predictor of L2 outcomes than storage-only measures. In line with this, studies employing more complex WM measures have demonstrated a significant role of WM in vocabulary learning through online reading. For instance, Teng (2024) employed the Reading Span Task (RST; Daneman and Carpenter 1980)—which captures both storage and executive processing—and found significant WM effects on incidental vocabulary learning through digital reading.

Yet whether WM plays a similarly important role in theme-based NR with AI-generated texts is less clear. On the one hand, reading across multiple texts requires cross-text vocabulary integration, which may place considerable demands on the central executive (Baddeley 2000)—suggesting WM may be particularly relevant in this context. On the other hand, the systematic repetition inherent in AI-generated texts may reduce processing burden, potentially diminishing the extent to which WM predicts learning outcomes. Indeed, if repetition is sufficiently massed and predictable, vocabulary gains may be driven largely by exposure itself, leaving little variance to be explained by executive resources—a possibility consistent with the null WM effects reported in

some computer-based reading studies (Varol and Erçetin 2016, Ünal 2023). The role of WM in this specific context has thus yet to be empirically examined.

Against this background, the present study investigates vocabulary learning through narrow reading with AI-generated texts and the extent to which individual differences in vocabulary size and WM—the latter reflecting learners' capacity to attend to and notice target lexis during reading—predict learning gains. Accordingly, the following research questions guide the study:

RQ1: To what extent does online narrow reading with AI-generated texts lead to vocabulary learning gains?

RQ2: To what extent do L2 vocabulary size and WM capacity predict vocabulary learning gains through online narrow reading with AI-generated texts?

3. Methods

3.1 Design and Context of the Study

The present study adopted a quasi-experimental pretest-posttest design with two groups: a narrow reading group and a comparison group. Participants were enrolled in four compulsory academic English classes for non-English majors at a university in South Korea, with two classes randomly assigned to the narrow reading group ($n = 48$) and the remaining two to the comparison group ($n = 48$). Because random assignment was implemented at the class rather than the individual level, the design involved intact classes. Each class met weekly for three hours, with the reading intervention implemented during the first hour of each session in a computer lab. Both groups participated in nine sessions over approximately ten weeks. Figure 1 illustrates the overall procedure of the study.

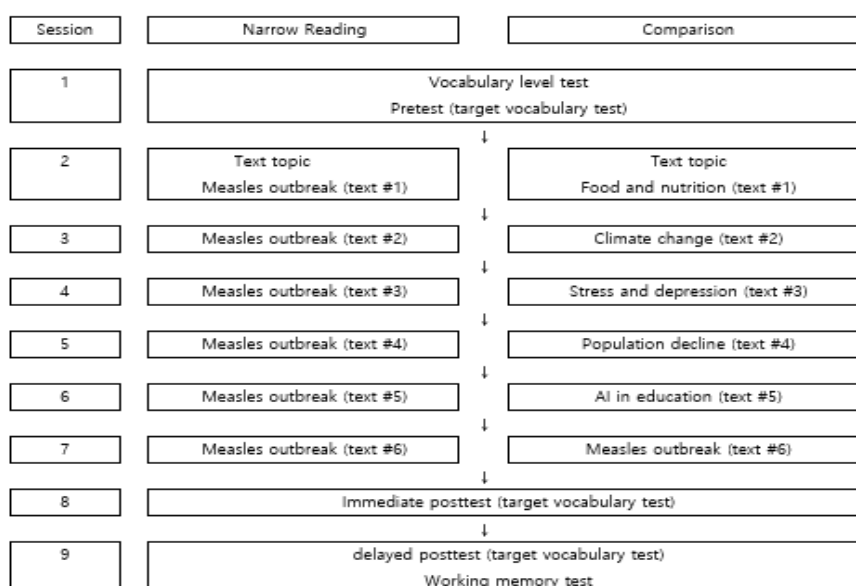


Figure 1. Procedure of the Study

In Session 1, all participants completed a vocabulary level test and a pretest of target vocabulary. Sessions 2 through 7 constituted the reading intervention, during which the narrow reading group read six AI-generated texts on a single topic (measles outbreak), while the comparison group read six AI-generated texts on six unrelated topics, one of which overlapped with the topic of the narrow reading group. Specifically, the comparison group's texts covered the following distinct topics: Food and Nutrition, Climate Change, Stress and Depression, Population Decline, AI in Education, and Measles Outbreak (Session 7). This design ensured that the comparison group received a single exposure to the target topic and its associated vocabulary, thereby reflecting a wide reading condition in which thematic repetition of target vocabulary across sessions was absent. Consequently, any observed differences in vocabulary gains between the two groups can be attributed to the differential effects of repeated, thematic exposure—the defining feature of narrow reading—rather than a complete absence of target word encounters in the comparison group. All texts were accessed online individually via computers in the lab. Session 8 included an immediate posttest of target vocabulary and a working memory test, and Session 9, administered two weeks later, included a delayed posttest.

3.2 Participants

A total of 96 first-year Korean EFL university students participated in the study, of whom 59 were male and 37 were female, with ages ranging from 18 to 21. Based on a practice TOEIC test administered at the beginning of the academic year, the large majority of participants fell within the B1–B2 bands of the CEFR, indicating intermediate to upper-intermediate proficiency. An independent samples *t*-test revealed no significant difference in initial English proficiency between the two groups ($t = 0.84$, $df = 94$, $p > .05$).

3.3 AI-Generated Reading Materials

Reading materials were prepared using Edcafe, an AI text-generation platform. The platform's “create reading activity” function was used to generate a reading text from a specified topic, together with a set of comprehension questions; generated texts could be edited to suit a target proficiency level. Figure 2 shows the interface used to generate the materials.

← Back

Tutorial

Create Reading Activity

Create reading activity from a topic, vocabulary list, text, files, or webpage.

Topic Text Webpage Files

Measles outbreak:causes, symptoms, and public health responses

Additional instructions (optional) Snippets

Include the following 8 words: outbreak, rash, uptake, contagious, measles, complication, epidemic, resurgence. The text level should be within the B1-B2 range. The text should be appropriate 600 words in length.

Output language

English

Generate Reading Activity

Figure 2. Screenshot of the “Create Reading Activity” Function in Edcafe AI

Figure 2 illustrates how the reading materials were generated. The topic was entered in the Topic field, and target vocabulary and text specifications were provided in the Additional Instructions field. Clicking the “Generate Reading Activity” button then produces a text tailored to the given instructions (Figure A1 in the Appendix). Edcafe AI also automatically generates multiple-choice comprehension questions for each text, with ten questions available per activity. A full sample text, along with comprehension questions and an AI-generated audio narration, is also available at <https://app.edcafe.ai/share/reading/measles-1-revised-69dba7d17af99efb06b5a252>.

All 11 texts used in the study were generated following this procedure, with topics presented in Figure 1. All texts were designed to fall within the CEFR B1–B2 range and to be approximately 600 words in length. For the narrow reading group, each text was created with instructions to include eight target words (see below for the target word selection procedure). For the comparison group, each text was instructed to include eight topic-relevant vocabulary items selected from beyond the 3,000 most frequent word families to ensure comparability with the target word levels. Specifically, for the narrow reading group, the prompt entered into Edcafe AI specified the topic as “Measles outbreak: causes, symptoms, and public health responses” and instructed the platform to include all eight target words at the B1–B2 level in approximately 600 words. For the comparison group, each prompt specified the relevant topic and instructed the platform to include vocabulary items from beyond the 3,000 most frequent word families at the same proficiency level and length. Following generation, all texts were reviewed by the researchers to confirm that the required vocabulary items were present and to verify linguistic accuracy, thematic coherence, and level appropriateness before use in the study. Post-generation editing was minimal: revisions were limited to replacing a small number of off-level words, adjusting occasional sentences for accuracy and naturalness, and correcting minor formatting issues; no structural rewriting of the texts was undertaken. Across the dataset, the eleven texts contained an average of 563 words (narrow reading texts: $n = 6$, $M = 565$; wide reading texts: $n = 5$, $M = 560$). The texts had an average Flesch–Kincaid Grade Level of approximately 9.1, indicating that a U.S. Grade 9 level of education is required to comprehend the texts.

3.4 Target Words

The selection of target words for the study proceeded as follows. First, ten newspaper articles on measles outbreaks were collected from online sources. Three instructors teaching the participants then identified words that were likely unfamiliar to the students but essential for understanding the topic. Through discussion, a final set of eight target words was established. To ensure that participants did not have prior knowledge of these words, the list was piloted with 21 students from similar backgrounds at the same institution using the same vocabulary test employed in the main study. The results indicated that the target words were largely unfamiliar to the pilot participants. All eight target words fell beyond the 3,000 most frequent word families: *outbreak*, *rash*, *uptake*, *contagious*, *measles*, *complication*, *epidemic*, and *resurgence*. This relatively small set of target words is consistent with prior narrow reading research (Dang and Lu 2025, Kang 2015) and was adopted to allow tight control over the frequency and distribution of each item across the six texts. Moreover, thematically core words that recur naturally across texts on a single topic are inherently limited in number, and embedding a larger set of low-frequency items in every text would have compromised the readability and naturalness of the materials.

3.5 Measurement of Vocabulary Size

Participants’ vocabulary size was assessed using the Vocabulary Size Test (VST; Beglar 2010). The test consists of 140 multiple-choice items, with 10 items sampled from each 1,000-word frequency band. This 14-level structure

allows for a more fine-grained estimation of vocabulary size across different frequency levels. Each item is presented in a sentence and followed by five response options—four Korean translations and an “I don’t know” option. Participants selected the option that best matched the meaning of the target word in context. In the present sample, the VST showed high internal consistency (Cronbach’s $\alpha = .89$).

3.6 Measurement of Working Memory

Working memory was assessed using a computerized version of the Reading Span Test (RST), administered in participants’ second language (L2). This choice was motivated by prior research indicating that complex span tasks, particularly when conducted in the L2, are more sensitive to individual differences in L2 reading comprehension (Shin 2020). The task was implemented using *Paradigm* (Perception Research Systems Inc.; Tagliaferri 2005), and the materials were adapted from Shin (2020) to match the intermediate proficiency level of the participants. The RST consisted of 70 sentences organized into sets ranging from two to eight sentences. All sentences were constructed using high-frequency vocabulary (1K–2K levels) and contained 10–12 words to minimize lexical confounds.

During the task, participants read each sentence silently and judged its semantic plausibility by pressing a “yes” or “no” key. Half of the sentences were plausible, and half were implausible. This judgment task ensured semantic processing while participants encoded the final word of each sentence for subsequent recall. After each set, participants recalled the sentence-final words in the correct order. Reading span scores were calculated following Leiser (2007) as a composite of three components: (a) reaction time for accurate judgments, (b) accuracy of semantic judgments, and (c) recall accuracy of sentence-final words. All components were standardized (z-scores), with reaction time reverse-coded, and averaged to yield a single composite score, where higher values indicated greater working memory capacity. Internal consistency, estimated from set-level recall scores, was acceptable ($\alpha = .81$).

3.7 Vocabulary Tests

Participants’ knowledge of the target vocabulary was assessed at two levels: form recognition and meaning recall. The form recognition test measured participants’ ability to identify the correct spelling of each target word using a multiple-choice format (Pellicer-Sánchez and Schmitt 2010). Each item included one correct answer, three distractors, and a “Don’t know” option. The distractors were designed to be morphologically and phonologically similar to the target words to enhance test difficulty and validity (see Table 1). Responses were scored dichotomously (1 = correct, 0 = incorrect).

Table 1. Sample Items for the Target Vocabulary Test

	Form recognition test	Meaning recall test
Sample item	Choose the correct spelling of the word from the options below, or select “Don’t know” if you are unsure. A. measels B. meesles C. measles (answer) D. mesales E. Don’t know	Provide the meaning of the word in either Korean or English. Measles: _____

The meaning recall test required participants to provide definitions in either English or Korean. Responses were scored on a 3-point scale (0, 0.5, 1). A score of 0.5 was assigned when the response conveyed the core meaning (e.g., synonym or paraphrase) but demonstrated partial knowledge, such as errors in part of speech. Both tests were administered as pretests and posttests using the same target items and 15 distractors, with item order randomized to minimize testing effects. All meaning recall responses were independently scored by two raters, with interrater agreement of 95.4% and 96% for the immediate and delayed posttests, respectively. All discrepancies were resolved through discussion, and the agreed-upon scores replaced the original ratings. The target vocabulary test demonstrated adequate reliability (Cronbach's $\alpha = .78$).

3.8 Data Collection Procedures

Data collection was conducted over nine sessions during regular class hours. In Session 1, participants completed an informed consent form, a background questionnaire, a vocabulary pretest, and the Vocabulary Size Test (VST). The reading intervention was implemented across Sessions 2–7.

In each session, participants accessed AI-generated texts and reading comprehension questions individually via computer through the course LMS. To ensure consistency in formatting and presentation across sessions and participants, all materials were exported to Word files and uploaded to the LMS rather than accessed via external web links. In addition, target words were not visually highlighted (e.g., bolded) to prevent unintended attentional bias. The comprehension questions accompanying each text were included to encourage engagement with the reading materials. On average, participants answered 80% of the comprehension questions correctly, indicating that they adequately understood the texts; these responses were not analyzed further.

Participants submitted all responses through the LMS. In the session immediately following the intervention (Session 8), participants took the immediate vocabulary posttest and the working memory test. Two weeks later, all participants completed the same vocabulary test again as a delayed posttest.

3.9 Data Analysis

Vocabulary test scores were analyzed using a repeated measures analysis of covariance (ANCOVA), with time (posttest and delayed posttest) as a within-subject factor, group (narrow reading vs. comparison) as a between-subject factor, and pretest scores as a covariate. Assumptions of ANCOVA, including homogeneity of regression slopes, normality of residuals, and homoscedasticity, were examined and met, supporting the appropriateness of the analysis. The repeated measures ANCOVA was adopted because the design involved a single between-subjects factor, only two posttest occasions, and one baseline covariate, and because this approach maintains comparability with prior NR research (e.g., Dang and Lu 2025).

To further examine the extent to which learners' L2 vocabulary size and working memory predicted vocabulary gains, Pearson correlation analyses were first conducted, followed by multiple regression analyses ($\alpha = .05$). In each regression model, vocabulary size and working memory were entered as independent variables, and vocabulary gain scores were treated as the dependent variable. As RQ2 specifically concerns vocabulary learning through narrow reading, regression analyses were conducted within the NR group only ($n = 48$). The comparison group was excluded from these analyses as participants did not engage in narrow reading. Two gain scores were calculated: (a) the difference between the immediate posttest and the pretest, and (b) the difference between the delayed posttest and the pretest. Preliminary analyses confirmed that the assumptions of linearity, normality, homoscedasticity, independence, and absence of multicollinearity were satisfied.

4. Results

4.1 The Effects of NR on Vocabulary Learning

As shown in Table 2, the eight target words recurred across the six narrow reading texts, yielding 115 target-word tokens in total (an average of 19.2 per text). Descriptive statistics for vocabulary test scores are presented in Table 3. The mean pretest scores were low for both groups, indicating limited prior knowledge of the target words. The NR group ($M = 1.42$, $SD = 0.88$) and the comparison group ($M = 1.39$, $SD = 0.56$) did not differ significantly at baseline, $t(94) = -0.21$, $p = .84$, confirming pretest equivalence between the two groups.

Table 2. Frequency of Target Word Occurrences across the Six Narrow Reading Texts

Target word	Text 1	Text 2	Text 3	Text 4	Text 5	Text 6	Total
<i>outbreak</i>	2	3	4	3	1	3	16
<i>rash</i>	2	3	1	1	2	2	11
<i>uptake</i>	2	3	2	2	1	2	12
<i>contagious</i>	1	2	1	2	1	2	9
<i>measles</i>	5	6	9	8	8	7	43
<i>complication</i>	1	2	1	1	2	2	9
<i>epidemic</i>	1	1	2	1	1	1	7
<i>resurgence</i>	1	1	2	1	1	2	8
Total	15	21	22	19	17	21	115

Note. Counts include inflected forms of each target word (e.g., outbreak/outbreaks, complication/complications).

Both groups demonstrated vocabulary gains from pretest to the immediate and delayed posttests; however, the NR group consistently outperformed the comparison group at both time points (see Table 3).

To examine group differences in vocabulary gains over time while controlling for pretest scores, a repeated measures ANCOVA was conducted with time (immediate posttest, delayed posttest) as a within-subjects factor, group (NR vs. comparison) as a between-subjects factor, and pretest scores as a covariate. The analysis revealed a significant main effect of the covariate, $F(1, 93) = 64.54$, $p < .001$, partial $\eta^2 = .410$, indicating that pretest vocabulary knowledge significantly predicted posttest performance. A significant main effect of group was also found, $F(1, 93) = 41.46$, $p < .001$, partial $\eta^2 = .308$, with the NR group ($M_{adj} = 4.88$) outperforming the comparison group ($M_{adj} = 3.42$) across both time points after controlling for pretest scores. A significant main effect of time was observed, $F(1, 94) = 17.90$, $p < .001$, partial $\eta^2 = .160$, indicating that scores decreased from the immediate posttest to the delayed posttest across both groups. The Group $\times$ Time interaction was not statistically significant, $F(1, 94) = 0.52$, $p = .473$, partial $\eta^2 = .005$, suggesting that the pattern of change over time did not differ significantly between the two groups.

Table 3. Descriptive Statistics of the Target Vocabulary Test Scores

Group	Pre <i>M (SD)</i>	Immediate <i>M (SD)</i>	Delayed <i>M (SD)</i>
Comparison ($n = 48$)	1.39 (0.56)	3.90 (1.50)	2.91 (1.37)
NR ($n = 48$)	1.42 (0.88)	5.25 (2.40)	4.55 (1.52)

Note. M = mean; SD = standard deviation. Maximum possible score is 16.

4.2 Vocabulary Size and WM as Predictors of Vocabulary Gains in the NR Group

Descriptive statistics for the vocabulary size test scores (see Table 4) indicate that participants had an average vocabulary size of approximately 5,768 words ($M = 5767.50$, $SD = 1542.01$). The experimental and comparison groups did not differ significantly in vocabulary size, $t(94) = 0.55$, $p = .581$. Vocabulary size varied considerably among participants, ranging from 3,000 to 8,500 word families.

Table 4. Descriptive Statistics of the Vocabulary Size Test

Group	$M (SD)$
Comparison	5855 (1584.76)
NR	5680 (1509.67)
Total	5767 (1542.01)

Note. Maximum possible score is 14,000.

Next, the experimental group's working memory test scores were analyzed to investigate whether WM predicted vocabulary gains. WM scores were computed by averaging the z-scores of reaction time, semantic judgment, and final word recall. The scores ranged from -1.68 to 3.00 with a mean of 0.67 , a standard deviation of 1.29 .

Before performing the regression analyses, Pearson product-moment correlation analyses were carried out. The results revealed that vocabulary size and WM were not significantly correlated ($r = 0.28$, $p > .05$), indicating a small and nonsignificant association between the two predictors and thereby supporting the assumption of no multicollinearity in multiple regression analysis (see Table 5 for the full set of correlations). In line with Plonsky and Oswald's (2014) discipline-specific guidelines, correlation coefficients of 0.25 , 0.40 , and 0.60 were interpreted as small, medium, and large, respectively. As presented in Table 5, both vocabulary size and WM showed significant positive relationships with immediate and delayed score gains.

A multiple regression analysis was conducted with pre-to-immediate gain scores as the dependent variable and vocabulary size and WM as predictors (Table 6). The model was significant, $F(2, 45) = 10.20$, $p < .001$, explaining 31.2% of the variance ($R^2 = .312$, adjusted $R^2 = .281$). Both vocabulary size ($\beta = 0.264$, $p = .042$) and WM ($\beta = 0.452$, $p < .001$) were significant positive predictors, with WM making a larger contribution to explaining immediate gains.

A parallel regression analysis was performed with pre-to-delayed gain scores as the dependent variable (Table 7). The model was also significant, $F(2, 45) = 9.76$, $p < .001$, accounting for 30.3% of the variance ($R^2 = .303$, adjusted $R^2 = .272$). WM was a significant positive predictor ($\beta = 0.46$, $p < .001$), whereas vocabulary size did not make a significant contribution when WM was included in the model.

Table 5. Correlation Coefficients between Observed Variables

	Vocabulary size	WM	Immediate gains	Delayed gains
Vocabulary size	-	0.28	0.39*	0.33*
WM	-	-	0.45**	0.52**

* $p < .05$, ** $p < .01$

Table 6. Multiple Regression for Gains between the Pretest and the Immediate Posttest

Predictors	B	SE	β	t	p	Adjusted R^2
Vocabulary size	0.001	0.001	0.264	2.096	0.042	
WM	2.103	0.587	0.452	3.585	<.001	.281

Table 7. Multiple Regression for Gains between the Pretest and the Delayed Posttest

Predictors	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	Adjusted <i>R</i> ²
Vocabulary size	0.001	0.000	0.198	1.523	0.135	
WM	1.670	0.470	0.461	3.550	<.001	.272

5. Discussion

This study examined the effectiveness of NR using thematically related AI-generated texts in a digital environment, as well as the predictive roles of vocabulary size and WM. The results demonstrated that learners made significant vocabulary gains following the NR intervention, and the NR group's advantage over the comparison group was sustained over time. Both vocabulary size and WM predicted vocabulary gains within the NR group; however, the predictive effect of vocabulary size was no longer significant after a two-week delay. In contrast, WM consistently emerged as a stronger predictor than vocabulary size for both immediate and delayed vocabulary retention, underscoring its prominent role in sustaining vocabulary learning outcomes.

5.1 Vocabulary Learning Gains from Online Narrow Reading with AI-Generated Texts

The first research question explored the overall effects of theme-based NR with AI-generated texts on EFL learners' vocabulary learning. After six sessions of reading one AI-generated text per session, the NR group's immediate posttest score averaged 5.25 out of 16, corresponding to a mean gain of 23.9% of the maximum possible score. This gain exceeded the 20.75% meaning recall reported by Dang and Lu (2025), although it fell short of the 68% gain in Kang (2015). The higher gains in Kang (2015) likely reflect the vocabulary glossaries provided alongside texts, which facilitated more direct form–meaning encoding. In contrast, the present study achieved its gains without such intentional learning support. The advantage over Dang and Lu (2025) may reflect the thematic alignment between target words and reading content: whereas their target items were academic words unrelated to the reading theme, the present study's target words were closely tied to the measles outbreak topic, potentially enhancing retention through repeated, meaningful exposure.

This interpretation aligns with previous findings on thematic relevance in vocabulary learning. Thematic clustering has been recognized as an effective organizational principle for vocabulary instruction (McDonald and Reynolds 2023, Tinkham 1997), and the repeated exposure to thematically linked words such as *contagious*, *epidemic*, and *resurgence* likely facilitated integration of these items into a coherent lexical network, strengthening form–meaning associations over time.

A key contribution of this study lies in its use of AI-generated materials through Edcafe AI, which enabled precise control over vocabulary load, text length, syntactic complexity, and thematic focus simultaneously—features that are difficult to calibrate in authentic materials (Webb 2020). This approach extends the work of Kang (2025), who implemented NR using Google News, by demonstrating the feasibility of AI-generated alternatives that offer greater pedagogical control. It is important to note, however, that because both groups read AI-generated texts, the present findings speak to the effectiveness of theme-based NR implemented with AI-generated materials rather than to the effectiveness of AI-generated texts relative to authentic or teacher-prepared materials.

Exposure to multiple texts on the same topic may also have strengthened students' background knowledge through content repetition. The NR group read six texts all covering measles—its symptoms, diagnosis, prevention, and outbreak—resulting in substantial content overlap across sessions. This stands in contrast to the comparison

group, who read six texts on unrelated topics and did not benefit from such thematic recycling. Growing topic familiarity across sessions may have progressively eased reading comprehension, which in turn could have supported vocabulary learning (Pulido 2004). Because background knowledge and topic familiarity were not directly measured in this study, however, this account remains speculative and awaits direct examination in future research.

The frequent and controlled recurrence of target words further contributed to vocabulary gains. All eight target words fell beyond the 3,000-word frequency level and are rarely encountered in general reading; however, AI-generated texts ensured their systematic embedding across all six texts. Target words also appeared in varied morphological forms—for example, *epidemic/epidemiological*, and *resurgence/resurge/resurgent*—providing repeated, context-rich encounters with both base forms and derivatives. This deliberate morphological distribution represents a methodological advantage over authentic sources, where such systematic exposure cannot be guaranteed, and is consistent with frequency-based accounts of vocabulary acquisition (Robinson and Ellis 2008, Trofimovich et al. 2012). At the same time, it should be acknowledged that the frequency of individual target words varied considerably across the texts (e.g., measles occurred 43 times, whereas epidemic and resurgence occurred only 7 and 8 times, respectively; see Table 2). Accordingly, part of the observed gains may reflect item-specific exposure frequency rather than thematic recycling per se, and the group-level advantage reported here should be interpreted with this imbalance in mind. Notably, whereas most prior NR studies assessed vocabulary gains only at an immediate posttest (e.g., Dang and Lu 2025, Kang 2015), the present study additionally examined retention over a two-week delay. The non-significant Group $\times$ Time interaction suggests that the NR group's advantage over the comparison group remained stable across both time points, indicating that although scores in both groups declined over the two-week delay, the relative benefit of thematic narrow reading persisted.

5.2 Predictive Effects of Vocabulary Size and WM on Vocabulary Gains

The second research question examined whether vocabulary size and WM predicted vocabulary gains within the NR group. Vocabulary size was a significant positive predictor of immediate gains, suggesting that pre-existing lexical knowledge supported the uptake of new words by enabling better text comprehension and greater attentional resources for unfamiliar items (Miralpeix and Muñoz 2018, Webb and Chang 2015, Feng and Webb 2020, Teng 2024). However, this effect did not persist at the delayed posttest once WM was controlled for. The systematic repetition built into the AI-generated texts likely reduced the advantage of larger vocabularies, as all learners received structured, sufficient exposure to each target word regardless of their lexical background—consistent with the introduction's hypothesis that AI-generated repetition may attenuate the role of vocabulary size. The field-specific nature of the target words may have further neutralized this advantage, as general vocabulary knowledge does not always transfer to specialized terminology (Hyland and Tse 2007, Dang and Lu 2025).

In contrast, WM consistently predicted both immediate and delayed vocabulary gains, with a larger standardized coefficient than vocabulary size at both time points. This enduring contribution is consistent with a language awareness account: learners with greater executive control are better positioned to notice target words as they recur across texts and to sustain attention to their form and meaning, which in turn supports the consolidation of these items into long-term lexical knowledge. Notably, this finding holds despite the reduced cognitive demands of NR: while thematic repetition eases processing load, cross-text vocabulary integration still places meaningful demands on executive functioning, suggesting that WM capacity remains a robust predictor regardless of material design. This finding aligns with Teng (2024) and contrasts with studies using simple digit span measures (Varol and Erçetin 2016), underscoring the importance of complex span tasks such as the RST in capturing WM's role in

vocabulary learning. By requiring participants to simultaneously retain sentence-final words in memory while making semantic plausibility judgments, and to subsequently recall those words in order, the RST captured multiple dimensions of WM—including storage capacity, processing speed, and executive control—that collectively appear central to noticing, encoding, and retaining new lexical items across multiple texts (Révész 2012). Theoretically, these findings extend noticing-based accounts of vocabulary acquisition (Schmidt 1990) to input environments in which repetition is deliberately engineered: even when the input itself guarantees repeated encounters with target words, learner-internal attentional resources continue to govern how effectively those encounters are converted into durable lexical knowledge. Input design and cognitive individual differences thus appear complementary rather than compensatory—optimized input does not eliminate the role of the learner.

6. Conclusion

This study demonstrates that engaging in online theme-based narrow reading (NR) with AI-generated texts can lead to meaningful vocabulary gains. Specifically, reading thematically related AI-generated texts helped EFL learners acquire topic-specific and specialized vocabulary that is rarely encountered in general reading, primarily through repeated and controlled exposure to target terms across multiple texts. In doing so, the study also makes a methodological contribution by demonstrating the feasibility of AI-generated texts as NR materials—showing that a platform such as Edcafe AI can be used to produce thematically coherent, proficiency-appropriate reading materials that systematically embed target vocabulary at sufficient frequencies, addressing a persistent challenge in NR material design. The findings also underscore the important roles of vocabulary size and working memory in predicting vocabulary learning outcomes through NR. While vocabulary size appears to support the initial uptake of new words, WM plays a more enduring role by contributing to the consolidation and long-term retention of lexical items. Learners with higher WM capacity were better able to retain newly acquired vocabulary and retrieve it successfully after a two-week interval.

The results of this study have several pedagogical implications. First, given that WM emerged as a stronger predictor of vocabulary gains than vocabulary size—particularly at the delayed posttest—instructors should consider adopting strategies that reduce cognitive load during NR. Pre-teaching target vocabulary before reading, providing brief glossaries, or introducing the topic through a short activating activity may help learners with lower WM capacity allocate more attentional resources to vocabulary encoding, thereby supporting more equitable learning outcomes across WM ability levels.

Second, the finding that vocabulary size predicted only immediate but not delayed gains suggests that even learners with smaller vocabularies can achieve comparable long-term retention when NR materials systematically recycle target vocabulary. When AI-generated texts ensure sufficient and controlled exposure to target words, the relative disadvantage of a smaller vocabulary may thus be offset, making theme-based NR viable for mixed-proficiency classrooms.

The results of this study should, however, be interpreted in light of several limitations. First, a potential limitation concerns the naturalness of the reading materials. Although AI-generated texts are not authentic in the traditional sense—that is, they were not produced by human writers for genuine communicative purposes—it is worth noting that large language models such as those underlying Edcafe AI are trained on vast amounts of authentic human-produced language data. As a result, AI-generated texts are likely to reflect patterns of natural language use, including vocabulary collocations, discourse organization, and syntactic structures characteristic of real-world texts. In this respect, AI-generated materials may occupy an intermediate position between fully

authentic and artificially constructed texts, rather than being categorically distinct from authentic input. Nevertheless, whether learners perceive AI-generated NR materials as engaging and credible, and whether subtle discourse-level differences from authentic texts affect the quality of vocabulary encounters, remains an empirical question that future research should address. Relatedly, the present study did not systematically evaluate the quality or naturalness of the AI-generated texts, nor did it examine learners' perceptions of these materials; accordingly, claims about the pedagogical value of AI should be confined to the demonstrated feasibility of AI-generated texts as NR materials. It should also be noted that both the narrow reading and comparison groups read AI-generated texts; the present study therefore does not test the effectiveness of AI-generated texts against authentic or human-written materials. Its contribution regarding AI is thus methodological—demonstrating a principled procedure for developing NR materials—rather than constituting direct evidence for the educational effect of AI itself.

Second, several word-related factors that may influence vocabulary learning were not systematically examined, such as cognateness, the availability of contextual clues, part of speech, and lexical type (e.g., abstract nouns, verbs, adjectives), all of which may have affected learning outcomes. Future studies should control for such lexical characteristics to better understand their role in vocabulary acquisition through NR. In addition, because participants completed a pretest of the target items, their attention may have been partially drawn to these words during reading; the learning observed here may therefore not have been entirely incidental. Relatedly, the target words differed considerably in how often they appeared across the texts (e.g., measles occurred far more frequently than the other items; see Table 2), and these differences in exposure may have influenced item-level learning outcomes. The small number of target items also precluded item-level analyses of such word-related factors and may have constrained the variance of the gain scores used in the regression analyses, which future studies with a larger set of items could address.

Third, the present study focused exclusively on verbal WM. To advance theoretical understanding, future research should also consider the visuospatial sketchpad and its interaction with verbal WM, which would provide a more nuanced account of the cognitive processes involved in vocabulary learning through digital reading.

Fourth, although the assumptions of the repeated measures ANCOVA were met, linear mixed-effects models—which treat participants and items as random effects—are increasingly standard for repeated measures data in applied linguistics and would additionally accommodate item-level variation, such as the unequal exposure frequencies across target words noted above; future studies are encouraged to adopt such models.

Finally, because the intervention was assigned at the class level—with two intact classes per condition—while analyses were conducted at the individual level, the potential clustering of students within classes could not be statistically modeled. Although the two groups were equivalent at baseline in proficiency, pretest vocabulary knowledge, and vocabulary size, future studies with a larger number of clusters should employ multilevel modeling to account for possible class-level variance.

Despite these limitations, this study provides empirical evidence that online theme-based NR with AI-generated texts can effectively promote the acquisition of specialized vocabulary, with learning gains maintained over a two-week period. The results also reveal that while vocabulary size facilitates the initial uptake of new words, WM plays a more decisive and enduring role in long-term retention. More broadly, the study suggests that engineered input and learner cognition must be understood together: AI can guarantee the quantity and distribution of lexical input, but what learners ultimately take from that input depends on the attentional resources they bring to it—an insight relevant both to theories of attention-driven vocabulary learning and to the design of AI-assisted vocabulary instruction.

References

- Ahmed, S. and Z.-H. Han. 2018. Models of reading. In J. I. Lontas, ed., *The TESOL Encyclopedia of English Language Teaching*, 1-7. Hoboken, NJ: John Wiley and Sons.
- Baddeley, A. D. 2000. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences* 4(11), 417-423.
- Baddeley, A. 2003. Working memory: Looking back and looking forward. *Nature Reviews Neuroscience* 4(10), 829-839.
- Beglar, D. 2010. A Rasch-based validation of the Vocabulary Size Test. *Language Testing* 27(1), 101-118.
- Çelik, F., C. Yangın Ersanlı and G. Arslanbay. 2024. Does AI simplification of authentic blog texts improve reading comprehension, inferencing, and anxiety? A one-shot intervention in Turkish EFL context. *International Review of Research in Open and Distributed Learning* 25(3), 287-303.
- Daneman, M. and P. A. Carpenter. 1980. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19(4), 450-466.
- Dang, T. N. Y. and C. Lu. 2025. Learning academic vocabulary through reading online news. *International Review of Applied Linguistics in Language Teaching* 63(4), 2231-2257.
- de Leeuw, L., E. Segers and L. Verhoeven. 2014. Context, task, and reader effects in children's incidental word learning from text. *International Journal of Disability, Development and Education* 61(3), 275-287.
- Feng, Y. and S. Webb. 2020. Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition* 42(3), 499-523.
- Gardner, D. 2008. Vocabulary recycling in children's authentic reading materials: A corpus-based investigation of narrow reading. *Reading in a Foreign Language* 20(1), 92-122.
- Hyland, K. and P. Tse. 2007. Is there an "academic vocabulary"? *TESOL Quarterly* 41(2), 235-253.
- Kang, E. Y. 2015. Promoting L2 vocabulary learning through narrow reading. *RELC Journal* 46(2), 165-179.
- Kang, E. Y. 2025. Tapping into Google News for theme-based narrow reading: Enhancing EFL learners' vocabulary use in writing. *Computer Assisted Language Learning*. Advance online publication.
- Koraishi, O. 2023. Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education and Technology Journal* 3(1), 55-72.
- Krashen, S. 2004. The case for narrow reading. *Language Magazine* 3(5), 17-19.
- Laufer, B. and T. Aviad-Levitzky. 2017. What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition? *The Modern Language Journal* 101(4), 729-741.
- Lee, S., H. Choe, D. Zou and J. Jeon. 2026. Generative AI (GenAI) in the language classroom: A systematic review. *Interactive Learning Environments* 34(1), 335-359.
- Leeser, M. J. 2007. Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning* 57(2), 229-270.
- Li, S. 2023. Working memory and second language learning: A critical and synthetic review. In A. Godfroid and H. Hopp, eds., *The Routledge Handbook of Second Language Acquisition and Psycholinguistics*, 348-360. New York: Routledge.
- Linck, J. A., P. Osthus, J. T. Koeth and M. F. Bunting. 2014. Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin and Review* 21(4), 861-883.
- McDonald, J. A. and B. L. Reynolds. 2023. Learning semantic and thematic vocabulary clusters through embedded instruction: Effects on very young English learners' vocabulary acquisition and retention. *Applied Linguistics Review* 14(5), 1129-1156.

- Miralpeix, I. and C. Muñoz. 2018. Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching* 56(1), 1-24.
- Nation, I. S. P. 2013. *Learning Vocabulary in Another Language*, 2nd ed. Cambridge: Cambridge University Press.
- Pellicer-Sánchez, A. and N. Schmitt. 2010. Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language* 22(1), 31-55.
- Plonsky, L. and F. L. Oswald. 2014. How big is “big”? Interpreting effect sizes in L2 research. *Language Learning* 64(4), 878-912.
- Pulido, D. 2004. The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning* 54(3), 469-523.
- Révész, A. 2012. Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning* 62(1), 93-132.
- Robinson, P. 2003. Attention and memory during SLA. In C. J. Doughty and M. H. Long, eds., *The Handbook of Second Language Acquisition*, 631-678. Oxford: Blackwell.
- Robinson, P. and N. C. Ellis, eds. 2008. *Handbook of Cognitive Linguistics and Second Language Acquisition*. New York: Routledge.
- Schmidt, R. W. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11(2), 129-158.
- Shin, J. 2020. A meta-analysis of the relationship between working memory and L2 reading comprehension: Does task type matter? *Applied Psycholinguistics* 41(4), 873-900.
- Tagliaferri, B. 2005. Paradigm [Computer software]. Perception Research Systems. Available at <http://www.perceptionresearchsystems.com>
- Teng, M. F. 2024. Working memory and prior vocabulary knowledge in incidental vocabulary learning from listening, reading, reading-while-listening, and viewing captioned videos. *System* 124, 103381.
- Tinkham, T. 1997. The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research* 13(2), 138-163.
- Trofimovich, P., L. Collins, W. Cardoso, J. White and M. Horst. 2012. A frequency-based approach to L2 phonological learning: Teacher input and student output in an intensive ESL context. *TESOL Quarterly* 46(1), 176-186.
- Ünal, B. 2023. Glossing and incidental vocabulary learning in L2 reading: A cognitive load perspective. *International Review of Applied Linguistics in Language Teaching* 61(2), 601-629.
- Varol, B. and G. Erçetin. 2016. Effects of working memory and gloss type on L2 text comprehension and incidental vocabulary learning in computer-based reading. *Procedia - Social and Behavioral Sciences* 232, 759-768.
- Verspoor, M. and N. Schmitt. 2013. Language and the lexicon in SLA. In P. Robinson, ed., *The Routledge Encyclopedia of Second Language Acquisition*, 353-360. New York: Routledge.
- Webb, S. 2020. Incidental vocabulary learning. In S. Webb, ed., *The Routledge Handbook of Vocabulary Studies*, 225-239. London: Routledge.
- Webb, S. and A. C. S. Chang. 2015. How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition* 37(4), 651-675.

Examples in: English

Applicable Languages: English

Applicable Level: Tertiary

Appendix. Sample AI-generated Reading Material and Comprehension Questions

Reading passage

A small town woke one spring to worrying news: a local school reported an unexpected **outbreak** of a childhood illness. Parents felt uneasy because the disease had returned after many years without cases. Health workers arrived quickly to learn which children had symptoms and to teach families how to protect themselves. The scene at the clinic showed how fast an infection can spread when people are not aware or prepared. The illness began with a high fever and a cough. After a few days, many children developed a red, blotchy **rash** that started on the face and then moved down the body. Doctors noticed that the illness matched the classic signs of **measles**, a viral infection that once caused frequent school absences and serious health problems. Measles is very **contagious**, which means it easily moves from person to person. A single infected child can pass the virus to many others in a busy classroom or a crowded bus. Public health teams responded with several actions. First, they confirmed the diagnosis using simple clinical checks and lab tests. Then they launched an information campaign, explaining how the virus spreads and which symptoms to watch for. The teams advised parents to keep sick children at home, to isolate cases until they were no longer contagious, and to seek medical help if breathing became difficult or if a child seemed unusually weak.

Figure A1. Excerpts from AI-Generated Text

Quiz

1 What sign led doctors to identify the illness as measles in the town?

- ☒ A A red, blotchy rash starting on the face and moving down the body
- ☐ B A persistent cough without other symptoms
- ☐ C Setting up free immunization clinics
- ☐ D Increased school absences

Explanation

The correct answer points to the combination of symptoms clinicians observed. Other options are either single symptoms not specific to measles or actions taken by health teams rather than diagnostic signs.

1. The rash that began on the face and spread down the body is a classic measles sign.
2. A cough alone can occur with many illnesses.
3. Vaccine clinics are a response, not a diagnostic sign.
4. School absences are an outcome, not a specific clinical sign.

Figure A2. Sample AI-Generated Comprehension Questions Produced by Edcafe AI