

Evaluating the Impact of Bilingual, Trilingual, and Multilingual Definitions on Incidental Vocabulary Learning through Story Reading

Raju Dhuli

Faculty of Humanities, Sharda University Uzbekistan

<raju.dhuli@shardauniversity.uz>

Rajakumar Guduru

Indian Institute of Technology Bhubaneswar, India

<rajakumarguduru@iitbbs.ac.in>

Praveen Kumar Ambati

Aditya Degree College, Ongole, India

<ambatipraveenkumar96@gmail.com>

Anand Binod Singh

Faculty of Humanities, Sharda University Uzbekistan

<anandbinod.singh@shardauniversity.uz>

Padma Lamo

Indian Institute of Technology Bhubaneswar, India

<a22hs09004@iitbbs.ac.in>

Abstract

This study investigates the effects of multi-glosses on incidental vocabulary learning through story reading among secondary school learners. It examines how different glossing conditions influence vocabulary acquisition and cognitive load. The sample consisted of 200 Class X students from diverse linguistic backgrounds assigned to three experimental groups and one control group. Each experimental group read fifteen short stories containing ten target words each, totaling 150 words, supported by glosses in one of the three language conditions, while the control group read the same stories without glosses. Vocabulary retention was measured using immediate and delayed post-tests, and cognitive load was assessed through the NASA Task Load Index (NASA-TLX), which evaluates mental demand, effort, performance, frustration, and related dimensions. Statistical analysis such as repeated mixed-ANOVA was used to examine differences in vocabulary gains and cognitive load. Results showed that all gloss conditions significantly improved immediate and long-term vocabulary retention compared to the control group. The bilingual glosses supported

better immediate recall due to reduced cognitive load, whereas trilingual and multilingual glosses promoted stronger long-term retention through deeper processing. However, multilingual glosses imposed higher cognitive demands. The study integrates multilingual glossing and cognitive load assessment, offering pedagogical insights for vocabulary instruction in multilingual-classrooms.

Keywords: Multilingual glosses, incidental vocabulary learning, cognitive load, ESL learners and story reading

It is widely acknowledged that expanding vocabulary knowledge is essential for mastering a language, as text comprehension and language output depend heavily on lexical competence (Teng, 2015). Laufer and Ravenhorst-Kalovski (2010) define vocabulary as a core component required for clear and fluent communication. They emphasize that vocabulary knowledge significantly predicts performance in reading comprehension and foreign language proficiency tests (Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011). Learners themselves often recognize vocabulary limitations as their primary expressive challenge, which highlights its central role beyond grammatical accuracy (Barani & Seyyedrezaie, 2017).

Vocabulary knowledge influences multiple dimensions of language development, including decoding, word recognition, reading fluency, and comprehension. Therefore, it serves as a strong predictor of learners' ability to process and understand texts effectively (Brooks et al., 2023; Zeng et al., 2025).

Studies show that deeper vocabulary knowledge supports higher-order cognitive tasks such as inference-making and summarization in reading (Dagnaw, 2023; Tong & Tong, 2022). Importantly, vocabulary knowledge and reading comprehension maintain a reciprocal relationship. Strong vocabulary knowledge facilitates comprehension, while meaningful reading exposure promotes vocabulary growth (Adlof, 2019; Cervetti et al., 2023; Elleman et al., 2019; Silverman et al., 2020).

Reading is a powerful source of incidental vocabulary acquisition, particularly when learners encounter unfamiliar words repeatedly in meaningful contexts. However, when lexical coverage is insufficient, comprehension breaks down because cognitive resources are diverted to word-level decoding rather than higher-level meaning construction (Qian & Lin, 2019). Empirical evidence confirms that repeated exposure through reading strengthens form–meaning mapping, enhances lexical representations, and supports long-term retention (Webb et al., 2023).

Given the importance of reading-based vocabulary development, substantial research has examined instructional strategies that enhance vocabulary acquisition. Explicit vocabulary instruction—characterized by direct teaching of word meanings, forms, usage, and morphological features—has been shown to significantly improve both receptive and productive vocabulary knowledge, particularly for lower-proficiency learners who require foundational lexical support (Adnan, 2020; Jeon, 2024).

In addition to explicit instruction, incidental vocabulary learning through extensive reading and graded readers has received considerable attention. Studies demonstrate that sustained reading activities promote lexical growth, especially when target words appear multiple times in context (Wang, 2020; Webb et al., 2023). The effectiveness of reading-based learning is further enhanced when texts include lexical support such as glosses. Research indicates that bilingual glossing

reduces cognitive load, increases noticing of unfamiliar words, and improves vocabulary retention compared to conditions without glosses (Yeou & Zyad, 2024).

Beyond reading-based and instructional approaches, researchers have investigated additional strategies to strengthen vocabulary development. Mnemonic techniques—such as the keyword method, imagery association, and semantic mapping—facilitate deeper encoding by creating meaningful associations between new and existing knowledge, resulting in improved vocabulary recall and retention (Fernando, 2021; Tsai et al., 2021).

In addition, technological interventions also play an important role in vocabulary learning. Mobile-assisted learning, spaced-repetition systems, and multimedia platforms provide adaptive practice and multimodal input, often producing greater vocabulary gains than traditional memorization approaches (Alisoy & Sadigzade, 2025; Daiu & Mihali, 2025; Yandira et al., 2025).

Furthermore, collaborative and task-based learning activities encourage learners to use new vocabulary in communicative contexts, thereby enhancing productive lexical knowledge through negotiation of meaning and peer interaction (El-haq et al., 2025). Vocabulary learning strategy training also contributes to lexical development by promoting learner autonomy through skills such as dictionary use, word analysis, and self-monitoring, which improve retention and metacognitive awareness (Mehmood et al., 2025).

Even more importantly, multilingual vocabulary contexts do indeed support retention by activating multiple linguistic pathways and improving vocabulary acquisition through cross-linguistic associations. Supporting this, Gyllstad et al. (2020) presented preliminary findings suggesting that vocabulary learning conditions involving multiple languages (English, Swedish, and other languages) yielded higher gain scores.

Similarly, Grace (1998) demonstrated that sentence-level translations significantly improved both short-term and long-term retention of word meanings. Additionally, Gyllstad et al. (2023) specifically noted that conditions involving L1 translation equivalents yielded higher immediate posttest scores, with a small effect size. These results were further supported by Šifrar Kalan et al. (2025), who found that students with high multilingual profiles had better vocabulary knowledge, especially at higher proficiency levels. Because it activates multiple linguistic pathways and strengthens cross-linguistic associations, exposing students to the target vocabulary in multilingual contexts promotes retention.

The repetition and semantic reinforcement provided through definitions in several familiar languages help students consolidate meaning more deeply, thereby facilitating long-term retention even if productive knowledge is not immediately assessed. Therefore, this study aims to evaluate the impact of three distinct definition types—bilingual, trilingual, and multilingual—on the incidental vocabulary acquisition of high school students through reading short stories.

Literature Review

Incidental and Intentional Vocabulary Learning

Incidental and intentional vocabulary learning represent two fundamental approaches in second and foreign language acquisition research. Incidental vocabulary learning refers to the unplanned acquisition of lexical items that occurs while learners engage in meaning-focused activities such as reading or listening, where attention is directed primarily toward message comprehension rather than memorization (Ilyas & Istaryatiningtias, 2025; Reynolds, 2025; Shahiwala et al., 2024).

However, defining incidental learning strictly as learning without intention is theoretically problematic, since learners' attention and intention may fluctuate across contexts and tasks (Webb et al., 2023).

In contrast, intentional vocabulary learning involves conscious and deliberate efforts to learn specific lexical items through explicit strategies such as word lists, flashcards, explanations, and memorization (Ni'mah & Sholihah, 2022; Webb et al., 2023). Reynolds (2025) emphasizes that intentional learning is characterized by a clear intention to commit words to memory, while a substantial portion of L2 vocabulary knowledge is also acquired informally through learner-initiated exposure beyond the classroom. L. Zhang (2020) further argues that intensive reading supports the development of language skills and literacy by encouraging learners to focus on vocabulary, reading strategies, and techniques, thereby cultivating stronger reading habits. In such contexts, learners may encounter vocabulary incidentally while also engaging in intentional analysis of lexical and grammatical forms.

Empirical research supports the contribution of various input modes to incidental vocabulary acquisition. Jin and Webb (2020) found that listening to teacher talk facilitates incidental learning of single words and collocations, while L1 translation promotes immediate vocabulary gains, although the frequency of occurrence did not significantly predict learning outcomes. Similarly, Feng and Webb (2020) reported that incidental vocabulary gains occur through reading, listening, and audiovisual input, with retention over time and with no significant differences across input modalities.

Recent intervention studies also highlight the effectiveness of structured reading-based approaches. Soleimani et al. (2022) demonstrated that both assisted and unassisted repeated reading improve incidental vocabulary learning and learner self-efficacy, although assisted reading with audio support yields significantly greater gains. Likewise, Barwasser et al. (2024) showed that a multi-component storytelling intervention incorporating flashcards and motivational reinforcement significantly enhanced receptive, expressive, and incidental vocabulary knowledge among young L2 learners.

In addition, Chachar et al. (2023) assert that integrating intensive reading fosters skilled and autonomous readers while positively influencing attitudes toward reading in English. Overall, the literature suggests that while intentional learning often produces stronger short-term recall due to focused attention on lexical items, incidental exposure through reading and meaning-focused input enhances depth of knowledge, collocational awareness, and long-term consolidation (Schmitt, 2010).

Meaning and Significance of Glosses

Glosses are generally defined as translations or brief explanations of unfamiliar or technical words that support learners in understanding texts (Segler et al., 2002). They may range from a single-word explanation to structured marginal notes or extended interpretive commentary (Stenner, 2020) and can appear in textual, pictorial, auditory, or multimodal forms. In addition, In ESL contexts, glosses provide immediate lexical support, reduce interruption during reading, and limit reliance on dictionaries.

Effects of Glossing on Vocabulary Learning

Research confirms that different glossing conditions significantly influence vocabulary acquisition and reading comprehension. Multimodal glosses—integrating textual, visual, and auditory

support—are particularly effective because they promote deeper cognitive processing and stronger vocabulary development (Ouyang et al., 2020; Paudel, 2025). Meta-analytic evidence shows that glossed reading generally improves L2 vocabulary acquisition compared to non-glossed conditions, with multiple-choice and L1 glosses often producing stronger gains (Amirian et al., 2016). However, outcomes are moderated by test format, learner proficiency, and text comprehension, while gloss mode itself does not always yield significant differences (Mahdi et al., 2024; Yanagisawa et al., 2020).

Types of Glossing and Their Effectiveness

In-text glosses that clearly signal lexical explanations tend to generate stronger learning gains because learners can easily identify and process target words (Ouyang et al., 2020; Z. Zhang & Webb, 2019). Similarly, the language of glossing also matters, as L1 translation glosses generally outperform L2 definitions in promoting vocabulary retention and comprehension, although their effectiveness depends on learner proficiency and contextual integration (Kim et al., 2020; Mahdi et al., 2024; Yanagisawa et al., 2020).

Furthermore, word meanings are further reinforced through pictorial and multimodal support. Studies comparing text-only, picture-only, and combined gloss formats consistently show that integrated text–image glosses yield the highest vocabulary gains. Multimedia glosses that combine images, audio, and text enhance meaning construction and retention by engaging multiple cognitive channels simultaneously (Hasan & Vaez-Dalili, 2025; Paudel, 2025; Yun, 2011). Nevertheless, task design plays a crucial role in determining gloss effectiveness. Research indicates that learners may pay limited attention to glosses when tasks do not require their use, suggesting that glosses are most effective when meaningfully integrated into instructional activities (Jung & Révész, 2018).

Empirical Evidence and Pedagogical Implications

In line with this evidence, recent intervention studies further confirm the pedagogical value of glossing. Classroom-based and quasi-experimental research demonstrates that multimodal and interactive gloss formats significantly improve vocabulary knowledge and incidental learning, particularly in multimedia environments (Hasan & Vaez-Dalili, 2025; Paudel, 2025). Overall, evidence suggests that multimedia glosses produce moderate to strong effects on vocabulary acquisition, though their effectiveness depends on learner characteristics, task design, gloss format, and language context. Therefore, gloss-based instruction should be strategically implemented rather than applied uniformly (Mahdi et al., 2024; Wu et al., 2025).

However, despite these findings, existing research has not sufficiently investigated the impact of different language-type models of glossing, such as bilingual, trilingual, and multilingual glosses, on vocabulary acquisition and retention. Most studies have primarily focused on L1 or L2 glossing. Multimodal glosses—integrating textual, visual, and auditory, leaving different language-type models of glossing underexplored. Accordingly, this study aims to bridge this gap by examining how different language-based gloss models influence vocabulary learning outcomes in multilingual contexts and by providing empirical evidence on their comparative effectiveness.

Importance of Graded Readers

D. Hill (2008) defines graded readers as “books written for learners of English using limited vocabulary and syntax (p. 185).” In this regard, they serve as progressive steps that help ESL learners gradually transition to reading authentic texts intended for native speakers (Wang, 2020).

Furthermore, they are often adapted from classic literature or original fiction and nonfiction works, including genres such as fantasy stories and biographies.

Historical and Linguistic Foundations

From a historical perspective, the development and success of graded readers are often attributed to West's (1953) *General Service List*, which identified the 2,000 most frequently used words in English. These texts are designed to help students expand their vocabulary while also exposing them to increasingly complex syntactic and narrative structures. Based on this foundation, these structures range from simple present tense constructions to more complex grammatical forms, such as conditional clauses and clauses of concession, as well as from sequential storytelling to the use of flashbacks. In addition, they are issued by several publishers, including Oxford, Pearson, Longman, and Cambridge (e.g., 250, 400, 700, and 1000) (Day & Bamford, 2002).

Pedagogical Value and Language Development

In pedagogical terms, graded readers have gained increasing recognition as valuable instructional materials for English language learners because they systematically scaffold reading development by providing simplified texts that match learners' proficiency levels. As a result, research consistently shows that graded readers support not only reading comprehension but also broader linguistic development. For instance, Alghizzi and Elyas (2022) found that exposure to graded readers during the COVID-19 pandemic improved university students' reading comprehension, highlighting the role of graded materials in maintaining and enhancing reading skills under challenging learning conditions. Likewise, Ateek (2021) reported that graded readers positively influenced both reading fluency and vocabulary knowledge for low-intermediate learners, demonstrating their multifaceted contribution to language development. Moreover, learner perceptions also point to the pedagogical value of graded readers. In a qualitative study at an Indonesian university, students noted that graded readers offered appropriate levels, useful vocabulary, and supportive illustrations that enhanced engagement and comprehension. Consequently, learners perceive these texts as both accessible and motivating (Ananda et al., 2021).

Extensive Reading and Digital Access

In addition to learner feedback, more recent research has also emphasized the role of graded readers in supporting extensive reading habits and vocabulary acquisition in young learners and adult ESL students alike (Abdelateef, 2021). Thus, consistent exposure to graded texts contributes to broader reading gains beyond comprehension alone. Finally, online graded reading materials also support extensive reading by aligning lexical demand with learner proficiency and ensuring effective comprehension (Housel & Oranjian, 2021; Wen & Yu, 2025).

Reading Comprehension and Short Stories

Short stories, when carefully selected and effectively implemented, serve as valuable instructional resources that enhance vocabulary, reading comprehension, and grammar among intermediate ESL learners (García-Cañarte & Ocaña, 2024; Mustafa Kurdi, 2022). In this regard, research highlights their pedagogical value in promoting reading development by offering authentic yet accessible language input that supports deeper comprehension.

For example, Mustafa and Jawad (2024) examined the effectiveness of short stories in improving EFL students' reading comprehension. The findings revealed that short stories significantly

enhanced students' reading skills by making texts more engaging and easier to understand. Similarly, Ghafar and Zubaedah (2023) investigated the effectiveness of short stories in ESL/EFL instruction and found that appropriately selected and taught short stories enhance vocabulary development, reading comprehension, and learner motivation. As a result, short stories are considered valuable resources for language classrooms.

In addition, Pratiwi et al. (2020) emphasized that engaging narratives increase students' motivation to read and encourage greater effort in learning. Furthermore, stories present vocabulary in memorable and real-world contexts, which helps students remember and apply new words accurately (Boers, 2022). At the same time, the number of languages used in glosses may influence cognitive load. Processing multiple linguistic cues can either facilitate comprehension or increase mental effort. Therefore, the division of cognitive resources when learners engage with bilingual or multilingual glosses may affect vocabulary acquisition and long-term retention (Zarifi & Azizinezhad, 2020).

Despite this, although research on glossing and short stories has grown independently, limited studies have examined how different language-based gloss types interact with cognitive load during narrative reading. Moreover, little is known about how this interaction influences vocabulary acquisition and retention. Thus, by analysing the interaction between gloss type, cognitive load, and vocabulary retention, this study explores how vocabulary acquisition can be optimized in reading-based interventions. Accordingly, the study is guided by the following research questions:

RQ1: Do learners in experimental groups outperform learners in control group on overall vocabulary posttests?

RQ2: Does the number of languages (two, three, or more) in the glosses correlate with improved immediate and long-term vocabulary retention among learners?

RQ3: What is the impact of bilingual, trilingual, and multilingual glosses on cognitive load during the vocabulary acquisition process?

Methods

Participants

The study's samples consisted of 200 Class X students from three different high schools in the states of Andhra Pradesh and Odisha, India. The participants were evenly distributed into four groups, with 50 students in each group. To ensure comparability, participants were selected based on their performance on an English proficiency test administered prior to the study. Only those who demonstrated comparable levels of English language proficiency and who shared similar linguistic repertoires were included. As a result, the bilingual, trilingual, and multilingual groups were equivalent not only in overall proficiency but also in their language backgrounds, thereby reducing potential confounds related to linguistic diversity.

Moreover, all the participants belonged to the same age group (15-16 years old). These students are divided into four groups: Group 1 students know two languages, namely Telugu and English; Group 2 students know three languages, such as Odia, Hindi, and English; and Group 3 students know four languages, namely Telugu, English, Odia, and Hindi. In contrast, group 4 students belong to control group where the control group contained a comparable distribution of bilingual, trilingual, and multilingual students, thereby minimizing potential confounding effects of

linguistic background. Additionally, all participants have been learning English as a compulsory subject from primary school through high school.

With regard to the instructional design, each group received reading instruction under a specific definition condition—bilingual, trilingual, multilingual, or no definition—to examine the effects of these variations on vocabulary acquisition. The treatment was carried out over 15 sessions.

Finally, this study did not require formal ethical approval, as it involved minimal risk and non-invasive educational research. However, all of the school's institutional heads and parents of the students gave their informed consent, and confidentiality and anonymity were strictly maintained throughout the study.

Research Design

The study employed a quasi-experimental design, in which vocabulary learning served as the dependent variable and the type of vocabulary definition functioned as the independent variable. In other words, definition types—bilingual, trilingual, multilingual, and without a definition—were chosen as the independent variable; on the other hand, the participants' scores on the immediate and delayed vocabulary post-tests served as the dependent variable in the study. For three months in the academic year 2024, the students participated in the experiment by reading fifteen short stories in each of the following four conditions: multilingual, trilingual, bilingual, and no definition.

Pre-Test and Group Equivalence

A pre-test was conducted in which students read a set of short graded stories similar in level and structure to the intervention materials. Students reported that the stories were generally easy to understand but noted difficulty with several unfamiliar words. This confirmed that reading comprehension levels across groups were comparable, while also demonstrating the presence of unknown vocabulary.

Since participants possessed different L1 backgrounds, and because meanings were provided in languages familiar to each group, it is reasonable to expect that L1-related factors influenced incidental vocabulary learning through glosses.

Treatment Procedure

In this study, participants assigned to the three definition conditions were designated as the experimental groups, while those in the no-definition condition comprised the control group. The control group received the same stories, reading time, and number of sessions as the experimental groups. However, their stories did not include any glosses or vocabulary support. No additional explanations, translations, or cues were provided. This ensured that the only difference between groups was the presence or absence of glosses, allowing vocabulary gains to be attributed specifically to glossing conditions. Following the reading sessions, all participants were instructed to complete the post-tests immediately after the short stories were read. Two weeks later, one by one, the delayed assessments were administered.

Measurement Focus

The assessments measured only receptive vocabulary knowledge—that is, students' ability to recognize and understand the meanings of words rather than produce them independently—this focus is still highly valuable for promoting long-term retention. This is because receptive

knowledge typically develops earlier and more reliably than productive knowledge, forming the foundation upon which productive skills are gradually built.

When learners repeatedly encounter target words in meaningful reading contexts, they strengthen semantic associations and deepen their understanding of how the words function in discourse. By assessing receptive knowledge, the study could determine whether students had successfully internalized the meanings of the target words well enough to recognize them during reading, which is a crucial first step toward productive mastery and use.

Materials and Instruments

The five instruments employed in this study included were Vocabulary Knowledge Scale (VKS), a set of fifteen graded-reader short stories from *Stories for Reproduction (Intermediate)* by L. A. Hill (1965), immediate and delayed post-tests, and the NASA-TLX scale; all of which were administered to every participant. These instruments are described as follows:

Vocabulary Knowledge Scale

Wesche and Paribakht (1996) developed the Vocabulary Knowledge Scale (VKS), which was used in this study to assess learners' prior familiarity with the 200 candidate vocabulary items. In this context, the VKS was intentionally used only as a diagnostic tool to identify unfamiliar vocabulary items prior to the intervention. More importantly, the VKS was not used as a pre-test or outcome measure, and vocabulary gains were assessed exclusively through multiple-choice immediate and delayed post-tests.

In terms of structure, the VKS consists of five levels, ranging from complete unfamiliarity to the ability to use the word accurately in a sentence. For the purpose of this study, words rated at Level 2 ('I have seen this word before, but I do not think I know what it means') were classified as unfamiliar.

Subsequently, the researcher selected 200 challenging words from the short stories and incorporated them into the VKS test. The results showed that students identified 150 of the 200 words as unfamiliar. Consequently, these 150 unknown words were chosen for the intervention.

Following this selection, the meanings of the words were presented as glosses in the stated language and placed in each of the fifteen stories while reading (see Appendix A). Throughout the study, word-recognition tests were employed to track students' progress in learning the target vocabulary.

Graded-reader Short Stories

A pilot study was conducted with 60 participants to assess the readability and comprehensibility of the graded-reader short stories selected from Oxford Stories for reproduction (see Appendix A) for the intervention. In this phase, these students were asked to read the stories and underline unfamiliar words. They were also interviewed regarding the clarity, difficulty level, and overall comprehensibility of the texts. The results of the interview revealed that the stories were easy to read and understand, but there were some unknown/ difficult words that the students underlined while reading the texts.

Based on the results of the pilot study, the researcher administered the VKS test to 200 sample participants to assess their level of word awareness. For this purpose, the researcher selected 200 challenging words from the short stories and incorporated them into the VKS test. The results

showed that students identified 150 of the 200 words as unfamiliar. Consequently, these 150 unknown words were chosen for the intervention.

Following this, each short story was adapted into three glossed versions—bilingual, trilingual, and multilingual—and one non-glossed version. In all glossed conditions, the target words were bold-faced in the text. At the bottom of each page, the target items were listed in order of appearance, accompanied by their meanings in the corresponding gloss languages: English and Telugu (bilingual), English and Hindi and Odia (trilingual), and English, Hindi, Odia, and Telugu (multilingual). In the control version, the story appeared without any glosses or definitions.

All glosses consisted of full, context-appropriate definitions rather than simple one-word translations. This ensured precise comprehension of each target item. Moreover, glosses across the four languages (English, Hindi, Odia, Telugu) were semantically equivalent and carefully aligned to maintain uniformity of meaning. To avoid semantic ambiguity and maintain consistency, English was used as a constant glossing language in all circumstances, whereas L1/L2/L3 glosses varied based on the learners' linguistic profiles. This ensured that multilingual glosses accurately represented the real multilingual settings in which Indian students frequently work.

In the multilingual condition, definitions were presented sequentially in the order English → Hindi → Odia → Telugu, consistent across all stories. The short stories used in the intervention were of moderate length, allowing students to read and interpret them without fatigue while still encountering a sufficient number of unfamiliar lexical items to facilitate vocabulary learning.

Immediate Post Tests

Before conducting the intervention, pretests were administered to all the students. Following this, students were instructed to read the short stories presented with glosses in different languages. After reading the short stories, the participants' vocabulary knowledge was measured by the immediate post-test. This test was designed to assess recognition of the 150 target words.

In terms of timing, the immediate post-tests were administered immediately after the completion of each story reading session, and a delayed post-test was administered 15 days later. Each form of the test (see Appendix B) consisted of multiple-choice items, one for each target word, with four options (one correct meaning and three distractors).

To minimize memorization effects, the delayed post-test used the same items but with question order and answer choices rearranged. There were 10 target words in total for each test, and students could earn up to 10 points by answering correctly, totaling 150 points.

Delayed Post Tests

The participants completed the same set of fifteen post-tests individually as delayed post-tests, administered 15 days after the initial immediate post-test. In this phase, the participants' varying levels of long-term memory and vocabulary retention across linguistic backgrounds were assessed using the delayed post-test. On the other hand, the quick post-test assessed the students' target word knowledge that they had inadvertently acquired from reading the graded readers.

The assessment expert reviewed both versions of the test and revised several items to improve precision and reduce ambiguity. Importantly, students were not informed in advance that they would take either the immediate or delayed vocabulary tests in order to promote conditions for incidental learning rather than intentional memorization.

Moreover, to enhance the reliability of the immediate and delayed vocabulary post-tests, all distractors were selected based on semantic similarity, part-of-speech consistency, and potential for plausible confusion. A language assessment specialist reviewed items to ensure content validity. Moreover, the test was piloted with a separate group of 60 students of similar proficiency. Based on pilot feedback, items exhibiting ambiguity or excessive difficulty were revised.

NASA Task Load Index (NASA-TLX) Scale

Cognitive Load Theory (CLT) posits that the human cognitive system is characterized by a limited working memory and an unlimited long-term memory (Sweller, 2011). In line with this theoretical framework, the NASA Task Load Index (NASA-TLX) developed by Hart and Staveland (1988) was employed to assess the cognitive workload associated with vocabulary learning through the three definition types—bilingual, trilingual, and multilingual—embedded in short stories. Notably, the NASA-TLX is a widely recognized subjective workload assessment instrument that enables participants to rate their perceived cognitive load across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration.

Regarding administration, the NASA-TLX questionnaire was administered once, immediately after the delayed post-test. Participants were instructed to rate their perceived cognitive load based on their experience during the reading sessions throughout the intervention period, rather than evaluating each story individually. This procedure, therefore, minimized fatigue and ensured that the ratings reflecting the overall cognitive demands of the glossing conditions.

In terms of structure, the questionnaire (see Appendix C) was administered to participants following the delayed post-tests. It consisted of six items corresponding to these six dimensions, each presented on a five-point Likert scale. Participants indicated their perceived workload related to learning vocabulary through the different gloss types by selecting responses ranging from "not at all" to "extremely."

Data Collection Procedure

Four groups of students from the same class were selected to participate in this study. Specifically, one group served as the control group, while the remaining three functioned as experimental groups. To assess participants' prior vocabulary knowledge, a Vocabulary Knowledge Scale (VKS) test was administered using 200 words drawn from the fifteen short stories. The results showed that students were unfamiliar with 150 of these words; therefore, these items were selected as the target vocabulary for the instructional intervention.

Regarding the duration of the study, although the overall research process extended for three months, the instructional intervention itself lasted for five weeks, comprising fifteen sessions during which students attended three 60-minute sessions per week, totaling 15 instructional sessions.

In terms of assessment, after each of the fifteen reading sessions, immediate post-tests were given. A delayed post-test corresponding to each story was administered exactly two weeks after its immediate post-test. This rolling schedule ensured that every target word was assessed after a consistent two-week interval. Importantly, a two-week interval is widely used in vocabulary acquisition research because it allows natural memory decay to occur while minimizing interference from external learning experiences (Schmitt, 2010; Waring & Takaki, 2003; Webb, 2005).

With regard to treatment conditions, students read fifteen short stories, each embedded with one of the three glossing conditions (bilingual, trilingual, or multilingual), depending on their group assignment. In contrast, the control group read the same stories without glosses. All students read each story silently and individually during class time. Furthermore, no oral reading, teacher explanation, or peer discussion was permitted in order to maintain uniform experimental conditions and prevent collaborative learning effects. Glosses were provided only in printed form, and no additional explanation or support was given.

sequence to ensure consistency of vocabulary exposure across conditions, all groups read the same set of stories in the same. Students were not informed in advance about the immediate post-tests and the delayed post-tests to maintain conditions conducive to incidental vocabulary learning. Throughout the intervention, English was used consistently as the medium of instruction across all sessions.

Data Analysis Procedure

The Statistical Package for the Social Sciences (SPSS, Version 28) was used to conduct both descriptive and inferential statistical analyses. To address Research Questions 1 and 2, students' vocabulary performance on the immediate and delayed post-tests was analyzed using a repeated mixed ANOVA to determine whether the four groups differed significantly in their vocabulary gains. Both immediate and delayed post-test scores were analyzed to evaluate short-term learning immediately after the intervention and long-term retention two weeks later.

When it comes to descriptive statistics, means and standard deviations were computed for all groups to summarize vocabulary performance across both testing points. To answer Research Question 3, data from the NASA-TLX questionnaire were analyzed descriptively to assess the cognitive load associated with the different glossing conditions. Means and standard deviations for each of the six TLX dimensions were calculated to interpret learners' perceived mental workload during the vocabulary learning process.

Results

The descriptive statistics in Table 1 revealed clear differences among the four groups across the three testing points. All groups began at almost identical pretest levels ($M \approx 40\text{--}41$), confirming that the learners had comparable vocabulary knowledge before the intervention. At the immediate posttest, however, large differences emerged.

Table 1. Descriptive Statistics for Immediate and Delayed Posttests Across Groups

Group	Pretest		Immediate		Delayed		N
	Mean	SD	Mean	SD	Mean	SD	
Bilingual	40.86	6.21	122.90	8.79	106.70	8.43	50
Trilingual	40.64	6.85	117.02	10.94	123.90	9.00	50
Multilingual	41.04	6.36	112.04	12.13	127.88	29.25	50
Control	40.92	6.47	44.02	9.71	42.72	8.05	50

Among the experimental groups trilingual and multilingual—showed dramatic improvements, with mean scores rising from pretest to delayed posttest, whereas the Bilingual and Control group showed only minimal increases for the immediate post-test. This indicates that the provision of glosses in different language combinations greatly enhanced incidental vocabulary learning during story reading, while the absence of glossing support resulted in negligible gains. Notably, the bilingual group demonstrated the highest immediate improvement, suggesting that bilingual definitions were the easiest to process and allowed for the quickest access to meaning.

A different pattern emerged in the delayed posttest, which examined vocabulary retention. Here, the multilingual group achieved the highest mean score followed by the trilingual group, indicating that definitions provided in three or more languages promoted deeper and more durable learning. Both these groups improved further from the immediate to the delayed posttest, suggesting ongoing consolidation of vocabulary knowledge. In contrast, the Bilingual group showed a moderate decline, implying that although bilingual glosses facilitated rapid learning, they did not support long-term retention as effectively as the trilingual and multilingual conditions.

The control group's delayed posttest score returned close to its pretest level, showing that vocabulary learned without gloss support was not retained over time. Overall, the results suggest that while bilingual glosses produce the fastest short-term learning, trilingual and multilingual glosses generate the strongest long-term retention, with multilingual glossing showing the most powerful effects on sustained vocabulary development.

Table 2 presents the results of a mixed ANOVA which revealed a significant main effect of Time on vocabulary scores, indicating that learners' performance changed substantially across the three testing points—pretest, immediate posttest, and delayed posttest. This effect was very large, as reflected in the high F-value and substantial partial eta squared showing that time alone accounted for more than 82% of the variance in learners' scores. Importantly, there was also a significant Time $\times$ Group interaction demonstrating that the four groups followed different developmental trajectories across the testing phases.

Table 2. Results of Mixed Repeated-Measures ANOVA: Within-Subjects Effects

Effect	df	F	p	Partial η^2
Time	2.119	684.138	< .001	.823
Time $\times$ Group	4.238	188.372	< .001	.719
Error (Time)	311.489	—	—	—

Although all groups performed equally in the pretest, indicating no initial differences, the immediate posttest revealed a clear divergence: the bilingual group showed the greatest short-term improvement, outperforming both the trilingual and multilingual groups, while the control group demonstrated the least progress. However, this pattern shifted by the delayed posttest, where the multilingual group surpassed both the bilingual and trilingual groups, indicating stronger long-term retention.

The large effect size for the interaction confirms that nearly 72% of the variance in score changes over time was due to differences among the glossing conditions. Overall, the strong main effect of Time combined with the substantial interaction effect shows that vocabulary scores increased significantly across sessions, and that the type of glossing support shaped not only the magnitude of improvement but also the durability of learning over time.

The between-subjects analysis presented in Table 3 examined the overall differences among the four groups—bilingual, trilingual, multilingual, and control—across all testing points. The ANOVA revealed a significant effect of Group on average vocabulary scores, indicating that group membership accounted for a substantial portion of the variance in learners’ scores. This confirms that the type of glossing support had a strong overall impact on learners’ vocabulary performance.

Table 3. Results of Mixed ANOVA: Between-Subjects Effects

Effect	df	F	P	Partial η^2
Intercept	1	26,225.436	< .001	.994
Group	2	1571.730	< .001	.955
Error	147			

Specifically, the experimental groups (bilingual, trilingual, and multilingual) achieved considerably higher average scores than the control group, demonstrating that providing definitions in one or more languages significantly enhanced learning compared to no glossing support. Among the experimental groups, the bilingual group showed the highest immediate gains, while the multilingual group exhibited superior performance in terms of delayed retention, reflecting the differential effects of gloss type on short-term learning versus long-term consolidation. The very large effect size underscores that the group differences were not only statistically significant but also practically meaningful, highlighting the strong influence of glossing conditions on vocabulary development.

When this between-subjects result is interpreted alongside the descriptive statistics and the interaction effect, it becomes clear that the experimental groups—bilingual, trilingual, and multilingual—experienced significantly greater improvement and retention than the control group. Therefore, the meaningful outcome for this study lies not in the static group means but in the dynamic learning gains, which strongly support the conclusion that the experimental groups outperformed the control group across the posttests.

Table 4. Descriptive Statistics of the NASA-TLX Questionnaire

	Bilingual		Trilingual		Multilingual	
Variable	Mean	SD	Mean	SD	Mean	SD
Mental	3.14	1.212	3.22	1.148	4.48	.646
Physical	2.60	1.106	3.28	1.069	4.24	.980
Temporal	2.38	.901	4.00	1.049	4.34	.917
Performance	3.00	1.087	3.70	1.23	3.94	1.167
Effort	2.80	1.030	3.82	.825	4.48	.646
Frustration	1.86	.700	2.66	.939	3.02	.820

The data mentioned in Table 4 reveal distinct differences in perceived cognitive load across the bilingual, trilingual, and multilingual gloss conditions. In terms of mental demand, the multilingual group reported the highest mean score, indicating a considerably greater cognitive burden compared to the bilingual and groups.

Similarly, physical demand increased with the number of glossing languages, with scores rising from bilingual to 4.24 multilingual. Temporal demand, reflecting the sense of being rushed, was notably low in the bilingual group moderate in the trilingual group and highest in the multilingual group indicating that learners experienced greater time pressure as the number of languages introduced increased.

Interestingly, in the performance domain, learners in the multilingual group perceived themselves as more successful than those in the bilingual or trilingual groups, which may imply that, despite the higher mental load, multilingual exposure supported deeper learning and vocabulary retention.

However, effort scores also increased significantly from bilingual to multilingual, confirming that learners in multilingual contexts had to exert substantially more cognitive effort. Finally, frustration levels were lowest in the bilingual group, and increased steadily across trilingual and multilingual conditions, reflecting the added complexity and potential for confusion in environments with multiple language inputs.

Table 5. MANOVA Results for Group Differences Across Cognitive Load Measures

Effect	Wilks' Lambda	F	Hypothesis df	Error df	p	Partial η^2
Group	.196	29.763	12	284	< .001	.557

Table 5 presents the results of the multivariate analysis using Wilks' Lambda, which revealed a statistically significant effect of Group on the combined cognitive load measures indicating a large effect size. This result suggests that approximately 56% of the variance in the combined dependent variables—mental demand, physical demand, temporal demand, performance, effort, and frustration—was accounted for by differences among the groups.

In other words, group membership significantly influenced how learners perceived the workload and task demands during the vocabulary learning activity. These findings justify further examination of univariate results to identify which specific dimensions of cognitive load were significantly affected by the type of glossing support (bilingual, trilingual, or multilingual).

Table 6. ANOVA Results for Each Dependent Variable

Dependent Variable	F	df	p	Partial η^2
Mental	26.43	2, 147	< .001	.265
Physical	30.58	2, 147	< .001	.294
Temporal	59.72	2, 147	< .001	.448
Performance	8.80	2, 147	< .001	.107
Effort	49.74	2, 147	< .001	.404
Frustration	25.85	2, 147	< .001	.260

The univariate ANOVAs for each cognitive load variable in Table 6 revealed significant differences among the three groups across all six measures. For Mental Demand, the effect of Group was significant, indicating that approximately 27% of the variance in perceived mental

demand was explained by group membership. Similarly, significant group effects were observed for Physical Demand and Temporal Demand, suggesting that participants in different glossing conditions experienced differing levels of physical effort and time pressure, with the largest effect observed for temporal demand.

Significant differences were also found for Performance, Effort, and Frustration. These results indicate that the type of glossing support not only affected participants' perceptions of workload but also influenced the effort they invested, their sense of success, and the degree of frustration experienced during the task. Overall, the largest effect sizes were observed for Temporal Demand and Effort, highlighting that time pressure and cognitive investment were the most sensitive measures to group differences.

Discussion

The present study examined the impact of bilingual, trilingual, and multilingual glosses on incidental vocabulary learning through story reading, with particular attention to short-term gains, long-term retention, and cognitive load.

Effects of Glossed Reading on Vocabulary Learning

First, the findings confirm that glossed reading significantly enhances incidental vocabulary acquisition compared to non-glossed conditions. Learners who received bilingual, trilingual, or multilingual definitions substantially outperformed those who read without gloss support.

This result aligns with previous meta-analytic and experimental studies demonstrating the facilitative role of glosses in vocabulary learning (e.g., Mahdi et al., 2024; Yanagisawa et al., 2020). Consistent with earlier research, glosses appear to reduce processing interruptions and enable learners to maintain focus on meaning while simultaneously forming lexical associations.

However, the present study extends previous work by systematically comparing *language-based gloss models* (bilingual, trilingual, multilingual), an area that has received limited empirical attention. Most prior studies have focused primarily on L1 versus L2 glossing (Amirian et al., 2016) or on multimodal glosses combining text, audio, and visuals (Hasan & Vaez-Dalili, 2025; Ouyang et al., 2020). By contrast, the current study demonstrates that the *number of languages included in textual definitions* meaningfully shapes both immediate learning and retention outcomes.

Differential Effects Over Time

A notable pattern emerged in the differential effects of gloss types over time. Bilingual glosses were associated with stronger short-term gains, suggesting that providing meaning in two languages allows for rapid semantic access and efficient initial encoding.

This finding is consistent with research showing that L1-supported glosses promote faster form–meaning mapping and immediate recall (Kim et al., 2020). The relative simplicity of bilingual glosses likely reduced processing demands, enabling learners to allocate more cognitive resources to comprehension.

In contrast, trilingual and multilingual glosses demonstrated stronger long-term retention. This suggests that exposure to multiple linguistic representations may deepen semantic processing and strengthen memory consolidation.

From a cognitive perspective, multiple language cues may promote elaborative encoding and richer associative networks. This interpretation aligns with Cognitive Load Theory (Sweller, 2011) and depth-of-processing frameworks, which suggest that meaningful cognitive engagement supports long-term retention when working memory is not overloaded.

Thus, a key contribution of this study is the distinction between short-term acquisition and long-term retention across gloss types. While earlier research has examined immediate and delayed effects (Yanagisawa et al., 2020), few studies have shown that different gloss models produce distinct learning trajectories over time. The present findings suggest that bilingual glosses may be optimal for quick vocabulary gains, whereas multilingual glosses may better support sustained vocabulary development.

Cognitive Load Findings

The cognitive load findings further clarify these learning patterns. Learners in multilingual conditions reported greater mental demand, effort, and time pressure than those in bilingual conditions. This aligns with research indicating that processing multiple linguistic cues can increase working memory demands (Bonilla Carvajal, 2025; Shain et al., 2022). The results support the assumption that adding languages to glosses increases intrinsic and extraneous load due to the need to process parallel semantic representations. Importantly, higher cognitive load did not necessarily hinder learning outcomes.

Despite reporting greater workload, learners in multilingual conditions demonstrated superior delayed retention. This pattern aligns with research suggesting that increased cognitive engagement can promote deeper processing and stronger memory consolidation when it does not exceed learners' processing capacity. For instance, Bjork and Bjork (2011) proposed the concept of *desirable difficulties*, arguing that appropriately calibrated learning challenges enhance long-term retention.

Similarly, Barcroft (2002) demonstrated that tasks requiring greater semantic processing can strengthen vocabulary learning, provided attentional resources are not overloaded. Meta-analytic findings by Yanagisawa et al. (2020) further indicate that glosses promoting elaborative processing yield stronger delayed gains, supporting the interpretation that multilingual glosses may encourage richer lexical encoding. At the same time, elevated frustration and temporal demand in multilingual conditions underscore the importance of instructional calibration. According to Sweller's (2011) Cognitive Load Theory, excessive intrinsic or extraneous load can impair learning when working memory capacity is exceeded.

Empirical studies on glossing (e.g., Bonilla Carvajal, 2025) suggest that processing multiple language cues may divide attention, particularly among lower-proficiency learners. Likewise, Kalyuga and Plass (2025) emphasize that instructional support must be adapted to learners' prior knowledge to prevent cognitive overload. Therefore, gloss complexity should be aligned with learners' linguistic repertoire, proficiency level, and cognitive readiness to maximize benefits while minimizing unnecessary mental strain.

Role of Short Story Reading

The study also reinforces the pedagogical value of short story reading as a vehicle for incidental vocabulary development. Consistent with research showing that narrative contexts promote engagement and lexical retention (Boers, 2022; Mehmood et al., 2025), the use of graded short stories provided meaningful exposure to target words in context. The interaction between narrative reading and gloss support appears particularly powerful, as learners were able to integrate lexical information within coherent discourse structures rather than isolated lists.

Contributions of the Study

Beyond confirming the effectiveness of glossed reading for incidental vocabulary learning, the present study makes several important contributions to the existing literature.

First, it is among the few empirical investigations to directly compare bilingual, trilingual, and multilingual textual gloss models within a controlled quasi-experimental framework, thereby extending prior research that has largely focused on L1 versus L2 glossing or multimodal gloss formats.

Second, the findings demonstrate that the quantity of languages included in glosses influences not only overall vocabulary gains but also the temporal pattern of acquisition, distinguishing between short-term improvement and long-term retention. This nuanced perspective moves beyond simple effectiveness comparisons and highlights how gloss complexity shapes learning trajectories over time.

Third, by integrating cognitive load measurement through the NASA-TLX instrument with vocabulary performance outcomes, the study provides a more comprehensive understanding of how linguistic complexity interacts with working memory demands during reading.

Finally, the research is situated within an authentically multilingual educational context in India, offering ecologically valid insights into gloss-based interventions in real-world multilingual classrooms. Together, these contributions deepen theoretical understanding and provide practical guidance for designing cognitively informed glossing strategies in diverse language learning settings.

Limitations and Suggestions

Despite the valuable insights offered by this study, several limitations must be acknowledged. First, the study focused exclusively on Class X high school students from diverse regional backgrounds in India, which may limit the generalizability of the findings to younger learners, adult learners, or those in different educational or cultural settings. Second, the study involved only 150 target vocabulary words embedded in short stories; different results may emerge with varied text genres, longer reading materials, or content-specific vocabulary.

Another limitation lies in the self-reported nature of cognitive load data, which, while useful, is inherently subjective and may not fully capture the actual mental processes involved. Furthermore, although reading was silent and independently monitored, teacher or classroom environmental effects cannot be fully eliminated. The study assessed only receptive vocabulary knowledge. Productive vocabulary growth, which may follow different developmental trajectories, was not captured. The delayed post-test was administered after two weeks. A longer longitudinal follow-up would provide deeper insight into long-term retention patterns. The findings may not generalize beyond the specific age group and school settings studied.

In light of these limitations, future research could explore a broader range of learner populations, including university students and adult ESL learners, to assess the broader applicability of multilingual glossing strategies. It is also recommended to incorporate objective cognitive load measures, such as eye-tracking or reaction time tasks, to supplement self-reports.

Furthermore, longitudinal studies examining retention over more extended periods and exploring the integration of glosses in digital reading environments or mobile applications could provide more practical implications for technology-enhanced language learning. Lastly, a qualitative component, such as interviews or learner diaries, could offer deeper insights into learners' experiences and strategies while interacting with different gloss types.

Conclusion and Implications

This study investigated the effects of bilingual, trilingual, and multilingual glosses on incidental vocabulary acquisition and cognitive load among secondary school students. The findings clearly show that the number of languages embedded in glosses plays a significant role in shaping both the immediacy and durability of vocabulary learning. The bilingual glosses produced the strongest immediate gains, reflecting reduced processing effort and more efficient initial word recognition. In contrast, trilingual and multilingual glosses, although cognitively more demanding, led to superior long-term retention.

This pattern suggests that the additional linguistic input engaged learners in deeper semantic processing, thereby promoting more stable lexical consolidation. Cognitive load analyses further clarified these outcomes. Multilingual conditions imposed higher mental, temporal, and effort-based demands, yet learners in these groups also reported stronger perceptions of performance—a finding consistent with the concept of *desirable difficulty*. The increased effort required to process multiple linguistic cues appears to have facilitated deeper encoding, supporting stronger retention over time.

These insights carry important implications for language pedagogy in multilingual settings. First, teachers and curriculum designers should consider adopting a phased glossing approach, beginning with bilingual glosses to support early comprehension and transitioning to trilingual or multilingual glosses to enhance deeper learning and long-term retention.

Second, instructional materials should also be adapted to reflect learners' linguistic repertoires, especially in regions where multilingualism is the norm.

Third, incorporating multilingual glosses into digital reading platforms may offer scalable and personalized vocabulary support. Finally, the study highlights the importance of leveraging learners' existing linguistic resources to enhance vocabulary acquisition. By balancing cognitive load with meaningful learning opportunities, educators can design more effective reading interventions that support both immediate comprehension and sustained lexical growth.

Acknowledgements

I would like to express my sincere gratitude to the school principals of the three schools for their support and assistance throughout this intervention. Additionally, I am deeply grateful to the students who supported me in collecting the data for this study; their active participation provided me with valuable insights. Moreover, the researcher consulted ChatGPT for fine-tuning the language. So to say, ChatGPT was used only for language improvement but not for the content of

this study, meaning the content and idea of the research are original and brainchild of the researcher.

About the Authors

Raju Dhuli is an Assistant Professor of ELT in the Faculty of Humanities at Sharda University Uzbekistan. He holds PhD in ELT from IIT Bhubaneswar, India. Dr. Raju holds TEFL Certificate from University of Toronto, Canada. He has participated in conferences held in Barcelona, Spain, and Frankfurt, Germany. His research areas include vocabulary development, language skills, academic writing, linguistics. ORCID ID: 0000-0001-8909-5103

Rajakumar Guduru is serving as an Assistant Professor of English Language Teaching at the School of Humanities, Social Sciences, and Management at the IIT Bhubaneswar, India. He earned his MA, M.Phil, and PhD degrees in ELT from the EFLU, India. His research interests encompass developing critical vocabulary for ESL learners, SLA, teacher education and development, and communication skills. ORCID ID: 0000-0002-0928-1166

Praveen Kumar Ambati is an English lecturer at Aditya Degree college in Ongole, India. He holds MA from Central University of Andhra Pradesh and has qualified UGC NET in literature. His research areas include language skills, applied linguistics and literature. He has published research articles in international journals and presented in conferences internationally. ORCID ID: 0009-0002-8611-4369

G. Anand Binod Singh is the Dean and Professor at the Faculty of Humanities at Sharda University Uzbekistan. He has been teaching English Language and Literature to undergraduate and postgraduate students for the past 18 years. After working in India for a decade, he moved to Malawi. His interests include Indian Writing in English, Technical English, and Soft Skills. ORCID ID: 0009-0000-7627-2131

Padma Lamo is a PhD scholar in English Language Teaching at the School of Humanities, Social Sciences, and Management at the IIT Bhubaneswar, India. Her research aims to create an inclusive educational space for school and college learners. Areas of interests include global Englishes, accent study, relation between cognition and language processing in neurodivergent learners, educational psychology etc. ORCID ID: 0009-0001-3223-8799

To Cite this Article

Dhuli, R., Guduru, R., Ambati, P. K., Singh, G. A. B., & Lamo, P. (2026). Evaluating the impact of bilingual, trilingual, and multilingual definitions on incidental vocabulary learning through story reading. *Teaching English as a Second Language Electronic (TESL-EJ)*, 30(2). <https://doi.org/10.55593/ej.30118a5>

References

Abdelateef, A. M. (2021). Using graded readers for developing ESL preparatory stage pupils' writing skills. *21*(239), 29-61.

Adlof, S. M. (2019). Prologue to the forum: Vocabulary across the school grades. *Language, speech, and hearing services in schools*, 50(4), 461-465. https://doi.org/10.1044/2019_LSHSS-19-00007

- Adnan, A. H. M. (2020). The effects of explicit vocabulary instructions on secondary ESL students' vocabulary learning. *3L The Southeast Asian Journal of English Language Studies*, 26(2), 158 – 172 <http://doi.org/10.17576/3L-2020-2602-12>
- Alghizzi, T. M., & Elyas, T. (2022). The effect of graded-reading websites/applications on EFL undergraduates' reading comprehension during COVID-19 pandemic. *Electronics*, 11(11), 1751. <https://doi.org/10.3390/electronics11111751>
- Alisoy, H., & Sadigzade, Z. (2025). Comparative analysis of mobile applications and traditional methods in vocabulary acquisition. *Educational Process: International Journal*, 17, e2025322.
- Amirian, S. M. R., Zareian, G., & Nour, S. (2016). Comparing the effects of L1 and L2 definition on incidental vocabulary learning through listening to stories. *Journal of Language Teaching and Research*, 7(3), 604–610. <http://doi.org/10.17507/jltr.0703.24>
- Ananda, M., Nisa, R., & Safura, S. (2021). Students' perceptions toward the use of Grammarly in checking grammar in assignments. *ACCENTIA: Journal of English Language and Education*, 1(2), 72-77.
- Ateek, M. (2021). Extensive reading in an EFL classroom: Impact and learners' perceptions. *Eurasian Journal of Applied Linguistics*, 7(1), 109-131. <http://doi.org/10.32601/ejal.911195>
- Barani, G., & Seyyedrezaie, S. H. (2017). Language learning and vocabulary:(A review). *Journal of Humanities Insights*, 2(1), 193-196.
- Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language learning*, 52(2), 323-363. <https://doi.org/10.1111/0023-8333.00186>
- Barwasser, A., Urton, K., Knaak, T., & Grünke, M. (2024). Intentional and incidental vocabulary acquisition through multi-component storytelling: The case of German L2 primary school students. *Language Teaching Research*, 28(6), 2453-2486. <https://doi.org/10.1177/13621688211047962>
- Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 2(59-68), 56-64.
- Boers, F. (2022). Glossing and vocabulary learning. *Language Teaching*, 55(1), 1-23. <https://doi.org/10.1017/S0261444821000252>
- Bonilla Carvajal, C. A. (2025). Interlinear translations reduce cognitive load on EFL vocabulary acquisition. *Íkala, Revista de Lenguaje y Cultura*, 30(1). <https://doi.org/10.17533/udea.ikala.356253>
- Brooks, G., Clenton, J., & Fraser, S. (2023). Exploring the importance of vocabulary for English as an additional language learners' reading comprehension. In *EAL Research for the Classroom* (pp. 35-58). Routledge.
- Cervetti, G. N., Fitzgerald, M. S., Hiebert, E. H., & Hebert, M. (2023). Meta-analysis examining the impact of vocabulary instruction on vocabulary knowledge and skill. *Reading Psychology*, 44(6), 672–709. <https://doi.org/10.1080/02702711.2023.2179146>

- Chachar, Z. A., Ahmed, S. T., & Khurram, S. (2023). Investigating the impacts of intensive and extensive reading approaches on the reading attitudes of Pakistani pre-university ESL learners. *Journal of Educational Research and Social Sciences Review (JERSSR)*, 3(4), 50-64.
- Dagnaw, A. T. (2023). Revisiting the role of breadth and depth of vocabulary knowledge in reading comprehension. *Cogent Education*, 10(1), 2217345. <https://doi.org/10.1080/2331186X.2023.2217345>
- Daiu, P. S., & Mihali, K. (2025). Spaced repetition and virtual flashcards: Mobile-assisted language learning (MALL) for English vocabulary and grammar. *Pakistan Journal of Life & Social Sciences*, 23(2). <https://doi.org/10.57239/PJLSS-2025-23.2.00188>
- Day, R. R., & Bamford, J. (2002). Top ten principles for teaching extensive reading. *Reading in a Foreign Language*, 14(2), 136–141.
- El-haq, S. G., Taha, A., & Maqsoud, A. (2025). Enhancing primary school vocabulary through task-based learning: A study of effectiveness and challenges. *BSU-Journal of Pedagogy and Curriculum*, 4(7), 222-253.
- Elleman, A. M., Oslund, E. L., Griffin, N. M., & Myers, K. E. (2019). A review of middle school vocabulary interventions: Five research-based recommendations for practice. *Language, Speech & Hearing Services in Schools*, 50(4), 477–492. https://doi.org/10.1044/2019_LSHSS-VOIA-18-0145
- Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 42(3), 499-523. <https://doi.org/10.1017/S0272263119000494>
- Fernando, W. S. A. (2021). The potential of semantic mapping to develop the vocabulary of second language learners. *Journal of New Frontiers in Education and Social Sciences*, 1(1), 1-13.
- García-Cañarte, A., & Ocaña, M. (2024). The use of short stories to improve vocabulary: understanding the perceptions of EFL students. *Ciencia Latina Revista Científica Multidisciplinar*, 8(4), 1030-1044. https://doi.org/10.37811/cl_rcm.v8i4.12342
- Ghafar, Z. N., & Zubaedah, P. A. (2023). The impact of short stories for English as a second language and foreign language students: A literature review. *Journal of Social Science (JoSS)*, 2(7), 595-604. <https://doi.org/10.57185/joss.v2i7.84>
- Grace, C. A. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *The Modern Language Journal*, 82(4), 533-544. <https://doi.org/10.1111/j.1540-4781.1998.tb05541.x>
- Gyllstad, H., Sundqvist, P., Källkvist, M., & Sandlund, E. (2020). Vocabulary learning under three different language conditions in six multilingual L2 English classrooms in Sweden. In *EuroSLA 30, 2020: EuroSLA 30*.
- Gyllstad, H., Sundqvist, P., Sandlund, E., & Källkvist, M. (2023). Effects of word definitions on meaning recall: A multisite intervention in language-diverse second language English classrooms. *Language learning*, 73(2), 403-444. <https://doi.org/10.1111/lang.12527>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental*

- workload* (Advances in psychology, Vol. 52, pp. 139–183). North-Holland.
[https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hasan, H. J., & Vaez-Dalili, M. (2025). Effects of three types of glosses on Iraqi ESL learners' incidental vocabulary learning from audiovisual input. *Innovare Journal of Education*, 13(4), 11–18. <https://doi.org/10.22159/ijoe.2025v13i4.54987>
- Hill, D. (2008). Graded readers in English. *ELT Journal*, 62(2), 184–204.
<https://doi.org/10.1093/elt/ccn006>
- Hill, L. A. (1965). *Stories for reproduction (Intermediate)*. Oxford University Press.
- Housel, D. A., & Oranjian, N. (2021). A case study of the use of Newsela PRO among adult, emergent bi/multilingual learners in a precollege program. *Adult Learning*, 32(2), 59-69.
<https://doi.org/10.1177/1045159520959474>
- Ilyas, H. P., & Istaryatiningtias, I. (2025). Blending extensive and intensive reading to enhance critical reading skills in the ESL context. *JOLLT Journal of Languages and Language Teaching*, 13(4), 1859-1872. <https://doi.org/10.33394/jollt.v13i4.15441>
- Jeon, Y. (2024). *The influence of implicit and explicit learning through reading in a school environment: EFL adolescents' L2 vocabulary acquisition* (Doctoral dissertation, University of Oxford).
- Jin, Z., & Webb, S. (2020). Incidental vocabulary learning through listening to teacher talk. *The Modern Language Journal*, 104(3), 550-566. <https://doi.org/10.1111/modl.12661>
- Jung, J., & Révész, A. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, 40(4), 755–780. <https://doi.org/10.1017/S0272263118000165>
- Kalyuga, S., & Plass, J. L. (2025). *Rethinking cognitive load theory*. Oxford University Press.
- Kim, H. S., Lee, J. H., & Lee, H. (2020). The relative effects of L1 and L2 glosses on L2 learning: A meta-analysis. *Language Teaching Research*, Online First.
<https://doi.org/10.1177/1362168820981394>
- Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size, and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30. <https://doi.org/10.64152/10125/66648>
- Mahdi, H. S., Mohsen, M. A., & Almanea, M. (2024). Multimedia glosses and second language vocabulary learning: A second-round meta-analysis. *Acta Psychologica*, 248, 104341.
<https://doi.org/10.1016/j.actpsy.2024.104341>
- Mehmood, A., Mehmood, S., & ul Haq, Z. (2025). Vocabulary learning strategies and their effect on reading comprehension among EFL students. *Social Sciences Spectrum*, 4(1), 766-781.
<https://doi.org/10.71085/ss.04.01.437>
- Mustafa, G. K., & Jawad, H. F. (2024). Improving EFL students' reading comprehension skill through short stories. *Qalaai Zanist Scientific Journal*, 9(1), 1130-1165.
<https://doi.org/10.25212/lfu.qzj.9.1.39>

- Mustafa Kurdi, S. (2022). The advantages of teaching short stories in ESL classrooms: A critical evaluation. *International Journal of Social Sciences & Educational Studies*, 9(3). <https://doi.org/10.23918/ijsses.v9i3p93>
- Ni'mah, D., & Sholihah, F. A. (2022). Reading intensively: What do the students really need? *EduLite: Journal of English Education, Literature and Culture*, 7(1), 16–31. <https://doi.org/10.30659/e.7.1.16-31>
- Ouyang, J., Huang, L., & Jiang, J. (2020). The effects of glossing on incidental vocabulary learning during second language reading: Based on an eye-tracking study. *Journal of Research in Reading*, 43(4), 496–515. <https://doi.org/10.1111/1467-9817.12326>
- Paudel, P. (2025, January). Use of multimodal glosses in teaching English vocabulary for non-English specialised undergraduates in public university in Nepal. In *Frontiers in Education* (Vol. 10, p. 1443803). Frontiers Media SA. <https://doi.org/10.3389/feduc.2025.1443803>
- Pratiwi, D. I., Putri, J., & Suhadi, A. (2020). Short story as a media for motivating students' improvement in reading. *Premise: Journal of English Education and Applied Linguistics*, 9(1), 30-41.
- Qian, D. D., & Lin, L. H. (2019). The relationship between vocabulary knowledge and language proficiency. *The Routledge handbook of vocabulary studies*, 66-80.
- Reynolds, B. L. (2025). Informal and incidental second language vocabulary learning. *Education Sciences*, 15(12), 1606. <https://doi.org/10.3390/educsci15121606>
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.
- Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, 95(1), 26–43. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>
- Segler, T. M., Pain, H., & Sorace, A. (2002). Second language vocabulary acquisition and learning strategies in ICALL environments. *Computer Assisted Language Learning*, 15(4), 409-422. <https://doi.org/10.1076/call.15.4.409.8272>
- Shahiwala, S., Rahul, D. R., & Baker, J. R. (2024). Incidental vocabulary learning: A scientometric review. *Research Methods in Applied Linguistics*, 3(3), 100160. <https://doi.org/10.1016/j.rmal.2024.100160>
- Shain, C., Blank, I. A., Fedorenko, E., Gibson, E., & Schuler, W. (2022). Robust effects of working memory demand during naturalistic language comprehension in language-selective cortex. *Journal of Neuroscience*, 42(39), 7412-7430. <https://doi.org/10.1523/JNEUROSCI.1894-21.2022>
- Šifrar Kalan, M., Muñoz-Basols, J., Robles-García, P., Strawbridge, T., & Sánchez-Gutiérrez, C. (2025). The impact of multilingualism and proficiency on L2 vocabulary knowledge: contrasting high and low multilinguals. *International Journal of Multilingualism*, 22(2), 433-456. <https://doi.org/10.1080/14790718.2024.2314626>
- Silverman, R. D., Johnson, E., Keane, K., & Khanna, S. (2020). Beyond decoding: A meta-analysis of the effects of language comprehension interventions on K–5 students' language and

literacy outcomes. *Reading Research Quarterly*, 55(S1), S207–S233. <https://doi.org/10.1002/rrq.346>

Soleimani, H., Mohammaddokht, F., & Fathi, J. (2022). Exploring the effect of assisted repeated reading on incidental vocabulary learning and vocabulary learning self-efficacy in an ESL context. *Frontiers in psychology*, 13, 851812. <https://doi.org/10.3389/fpsyg.2022.851812>

Stenner, R. (2020). Gloss. In *Oxford Research Encyclopedia of Literature*.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>

Teng, F. (2015). Assessing the relationship between vocabulary learning strategy use and vocabulary knowledge. *PASAA: Journal of Language Teaching and Learning in Thailand*, 49(1), 39–65. <http://doi.org/10.58837/CHULA.PASAA.49.1.2>

Tong, X., & Tong, S. X. (2022). How vocabulary breadth and depth influence bilingual reading comprehension: Direct and indirect pathways. *Learning and Individual Differences*, 100, 102227. <https://doi.org/10.1016/j.lindif.2022.102227>

Tsai, M. H., Chang, L. Y., Chen, H. C., & Lin, C. L. (2021). Effects of key-image mnemonics on Chinese instruction for first-grade students' achievement and interest toward Chinese learning. *International Journal of Educational Research*, 109, 101856. <https://doi.org/10.1016/j.ijer.2021.101856>

Wang, Q. (2020). Incidental L2 vocabulary acquisition and reading: Concerns, progresses and future directions. *Chinese Journal of Applied Linguistics*, 43(4), 469-488. <https://doi.org/10.1515/CJAL-2020-0031>

Waring, R., & Takaki, M. (2003). At what rate do learners forget vocabulary? *Reading in a Foreign Language*, 15(2), 130–163.

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, 27(1), 33–52. <https://doi.org/10.1017/S0272263105050023>

Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, 56(2), 161-180. <https://doi.org/10.1017/S0261444822000507>

Wen, J., & Yu, H. (2025). The lexical profile of online graded reading materials in English language teaching: A corpus-based study. *Language Teaching Research*, 13621688251320465. <https://doi.org/10.1177/13621688251320465>

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review*, 53(1), 13-40. <https://doi.org/10.3138/cmlr.53.1.13>

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

Wu, Y., Liu, X., Zhao, Y., Chen, X., & Gao, J. (2025). Multimodal glosses revisited: Interplay of individual differences and repeated exposures. *Language Learning & Technology*, 29(30), 92-134. <https://doi.org/10.64152/10125/73649>

- Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading? A meta-regression analysis. *Studies in Second Language Acquisition*, 42(2), 411-438. <https://doi.org/10.1017/S0272263119000688>
- Yandira, Y., Marzuki, M., & Abrar, A. E. Y. (2025). Spaced repetition and multimedia encoding in digital vocabulary learning: Evaluating the effectiveness of memrise for adolescent EFL learners. *Global Englishes and Transformation Studies Journal*, 1(1), 57-68.
- Yeou, M., & Ziad, H. (2024). The impact of L1 and L2 glosses on EFL learners' incidental vocabulary learning through reading: A systematic literature review. *International Journal of Linguistics and Translation Studies*, 5(3), 81-95. <https://doi.org/10.36892/ijlts.v5i3.479>
- Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, 24(1), 39–58. <https://doi.org/10.1080/09588221.2010.523285>
- Zarifi, A., & Azizinezhad, Z. (2020). Vocabulary learning and retention: Cognitive load framework on trial. *Porta Linguarum: revista internacional de didáctica de las lenguas extranjeras*, (33), 265-278. <https://doi.org/10.30827/portalin.vi33.26660>
- Zeng, Y., Kuo, L. J., Chen, L., Lin, J. A., & Shen, H. (2025). Vocabulary instruction for English learners: A systematic review connecting theories, research, and practices. *Education Sciences*, 15(3), 262. <https://doi.org/10.3390/educsci15030262>
- Zhang, L. (2020). Teaching design and practice of intensive reading course based on BOPPPS. *Journal of Language Teaching and Research*, 11(3), 503–510. <https://doi.org/10.17507/jltr.1103.21>
- Zhang, Z., & Webb, S. (2019). The effects of reading bilingual books on vocabulary learning. *Reading in a Foreign Language*, 31(1), 108–139.

Appendices

Appendix-A: Stories with Multilingual Glosses

Johnny was four years old, and his favourite game was cowboys and Indians. He had a cowboy **suit** and a belt with two guns, and spent most of his time **pretending** to be fighting Indians. One day his mother took him in a train for the first time. Of course, he wore his cowboy **suit** and carried his two guns. He had seen a film of an attack by Red Indians on a train in the Wild West, so his mother was not surprised when he began playing at cowboys and Indians in the train. But when he wanted to open the window wide so that he could shoot out of it, she thought this too **dangerous**, and allowed him to have it open only at the top, so that he could shoot out of it if he stood up, but could not fall out. He was playing happily, **hiding** behind the curtain, **suddenly** stepping forward, firing a shot out of the window and then quickly stepping back again, when he **suddenly** gave a cry, fell back on to the seat, and lay there with his chin resting on his chest and his arms **hanging** loosely beside him. Of course, his mother was **frightened**. She thought that something from outside the train must have hit him as he stood at the window. She **shook** the child gently, but he made no movement, and his eyes **rolled** in his head. His mother was now very worried indeed. She picked Johnny up in her arms to go and find help; but just then he lifted one of his arms with great effort, pointed to his chest and said in a weak **whisper**, ‘Pull the arrow out, will you?’

1. Suit

- English: A piece of clothing worn for formal occasions
- Hindi: पोशाक
- Odia: ପୋଷାକ
- Telugu: వస్త్రం

2. Dangerous

- English: Likely to cause harm or injury.
- Hindi: खतरनाक
- Odia: ବିପଦଜନକ
- Telugu: హానికరమైన

3. Suddenly

- English: Quickly and without warning.
- Hindi: अचानक
- Odia: ହଠାତ୍
- Telugu: తక్షణమే

4. Hanging

- English: Drooping downward from the body in a loose way
- Hindi: लटकना
- Odia: ଝୁଲିବା
- Telugu: వేలాడుట
-

Appendix-B: Immediate Test (Multiple Choice)

1. Suit

The word *suit* most nearly means:

- a) A set of clothes worn together
- b) A formal jacket worn with trousers or a skirt
- c) A special outfit used for a particular activity (e.g., diving suit)
- d) A type of musical instrument

2. Dangerous

The word *dangerous* most nearly means:

- a) Likely to cause harm or injury
- b) Risky and unsafe
- c) Potentially harmful in certain situations
- d) Somewhat challenging but not harmful

3. Suddenly

The word *suddenly* most nearly means:

- a) Quickly and without warning
- b) All at once, in an unexpected way

- c) Abruptly, without preparation
- d) Gradually over a long time

4. Hanging

The word *hanging* most nearly means:

- a) Drooping downward from the body in a loose way
- b) Suspended from something above
- c) Positioned so that it can swing freely
- d) Fixed tightly against a surface

Appendix-C: NASA-TLX Questionnaire

1. How mentally demanding was the vocabulary learning task?
Not at all demanding () Slightly demanding () Moderately demanding ()
Very demanding () Extremely demanding ()
2. How physically demanding was the task (e.g., due to reading, understanding content)?
Not at all demanding () Slightly demanding () Moderately demanding ()
Very demanding () Extremely demanding ()
3. How hurried or rushed did you feel while completing the vocabulary task?
Not at all () Slightly () Moderately rushed () Very () Extremely ()
4. How successful do you think you were in learning the new vocabulary?
Not at all Successful () Slightly Successful () Moderately Successful () Very Successful ()
Extremely Successful ()
5. How much effort did you need to put in to understand and remember the vocabulary?
Very Little Effort () Little Effort () Moderate Effort () Much Effort () A Great Effort ()
6. How frustrated, irritated, or annoyed did you feel during the task?
Not at all () Slightly () Moderately rushed () Very () Extremely ()

Copyright of articles rests with the authors. Please cite TESL-EJ appropriately.