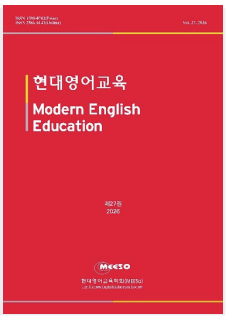
Modern English Education, Vol. 27, (2026) pp. 271-291
doi: https://doi.org/10.18095/meeso.2026.27.1.271



Modern English Education

ISSN: 1598-0782 / e-ISSN: 2586-6141
http://journal.meeso.or.kr

Deceptively Transparent Words and L2 Reading Comprehension among Korean High School EFL Learners

Doohwan Lee (Korea University)

Received: 1 April, 2026
Revised: 3 May 2026
Accepted: 15 May 2026

Lee, Doohwan. (2026). Deceptively transparent words and L2 reading comprehension among Korean high school EFL learners. *Modern English Education*, *27*, 271-291.

Keywords
Deceptively transparent words, L2 reading comprehension, Korean high school EFL learners, vocabulary breadth, vocabulary depth
기만적 투명어, 제2언어 읽기 이해, 한국 고등학교 EFL 학습자, 어휘 폭, 어휘 깊이

Doohwan Lee
PhD Candidate
Department of English Language Education
Korea University
Email: idu010255@korea.ac.kr

Abstract

This study examined reading comprehension related to deceptively transparent (DT) words—lexical items that seem familiar but can trigger contextually inappropriate meanings—among 40 Korean 11th-grade EFL learners. Employing a within-subject, counterbalanced Latin square design that partially varied task order and text topic across participants, the study included measures of vocabulary breadth, vocabulary depth, background knowledge, the DT sentence interpretation task, and reading comprehension for two texts on different but comparable topics: a control text without DT words and a DT text containing DT words. The results indicated that (1) mean reading comprehension scores did not significantly differ between the control text and the DT text, (2) performance on the DT sentence interpretation task did not explain additional variance in DT text reading comprehension, and (3) vocabulary depth uniquely accounted for variance in DT text reading comprehension beyond vocabulary breadth. Collectively, these findings suggest that understanding DT text may be more closely linked to individual differences in vocabulary depth rather than the mere presence of DT words, although this interpretation should be considered in the context of learners' overall reading comprehension levels and the specific conditions of the study.

INTRODUCTION

Reading comprehension is not simply the sum of isolated subskills; rather, it is a cognitive process that emerges from interactions among the reader, the text, and the reading activity (Snow, 2002). Because written texts provide only potential meaning, readers must actively construct an integrated interpretation by drawing on background knowledge, cultural knowledge, and their purposes for reading (Widdowson, 1979). Successful comprehension therefore depends on the coordinated contributions of multiple components, including decoding, vocabulary knowledge, syntactic knowledge, and higher-order cognitive skills (Alderson & Urquhart, 1984; Grabe & Stoller, 2011). Although no single component alone determines comprehension, vocabulary is particularly crucial because efficient lexical access supports the construction of sentence- and text-level meaning (Jeon & Yamashita, 2014; Laufer, 1997; Perfetti & Stafura, 2014; Qian, 2002).

Lexical difficulty, however, is not limited to unfamiliar forms. In second language (L2) reading, comprehension may

break down not only when readers encounter unknown words, but also when they recognize a word form yet assign it a context-inappropriate meaning (Laufer, 1989a, 1990, 1997; Perfetti, 2007; Perfetti & Hart, 2002; Perfetti & Stafura, 2014). In the present study, such items are termed deceptively transparent (DT) words, lexical items whose familiar form invites a plausible interpretation that does not match the meaning required by the context (Laufer, 1989a, 1990, 1997). In this sense, DT difficulty is treated not simply as unknown-word difficulty, but as a problem of context-sensitive meaning selection during reading.

This issue may be especially important for Korean high school EFL learners. In Korea, high-stakes English testing produces strong washback, and classroom instruction is often closely tied to exam preparation and multiple-choice reading tasks (I. Choi, 2008). Within this exam-oriented context, vocabulary breadth, often measured as vocabulary size, has been shown to relate to L2 reading comprehension among Korean EFL learners (M. Kim, 2019; J. Year, 2022). Although vocabulary breadth explains significant variance in reading comprehension (Jeon & Yamashita, 2014; Qian, 2002), it may be insufficient when learners encounter apparently familiar words whose context-appropriate meanings must be correctly selected. In such contexts, learners may fail to suppress a misleading interpretation, increasing the risk of local misunderstanding during reading.

Prior research on L2 reading comprehension has largely focused on broad predictors such as vocabulary knowledge, inferencing ability, and the proportion of unknown words (Hu & Nation, 2000; Jeon & Yamashita, 2014; Nassaji, 2003a, 2003b; Qian, 1999, 2002; Schmitt, 2008; Schmitt et al., 2011). Within this line of research, vocabulary breadth generally refers to the number of words learners know, whereas vocabulary depth is defined as how well those words are known (Nation, 2001; Qian, 1999, 2002; Read, 1993). Although this work has clarified major sources of variation in comprehension, less is known about how such factors operate when learners encounter DT words. Prior DT research, in turn, has not fully clarified whether DT-related meaning errors lead to problems in understanding connected texts. Much of the existing literature has focused on isolated lexical items, successful lexical inferencing, and translation performance rather than on how DT words affect cumulative understanding across clauses and sentences (Bensoussan & Laufer, 1984; Laufer, 1989a, 1990, 1997). This gap is theoretically important because reading comprehension depends on cumulative interpretation, and a local lexical error may distort subsequent propositions and weaken discourse-level integration (Kintsch, 1988). It also highlights the need for a more direct measure of learners' ability to interpret DT-containing sentences appropriately (Laufer, 1989a, 1990, 1997). In addition, it remains unclear how vocabulary breadth and vocabulary depth may differentially contribute to comprehension under DT conditions.

Against this backdrop, the present study examines whether DT words may represent a distinguishable lexical challenge in Korean high school EFL reading, rather than merely reflecting general limitations in vocabulary knowledge. Specifically, it investigates whether learners' reading comprehension differs between a control text condition and a DT text condition, whether performance on a DT sentence interpretation task accounts for additional variance in DT text reading comprehension beyond relevant task- and learner-related factors, and what the relative contributions of vocabulary breadth and vocabulary depth are to DT text reading comprehension. In doing so, the study seeks to clarify whether and to what extent DT-related difficulty can be distinguished from the effects of general reading ability, background knowledge, and vocabulary knowledge more broadly.

LITERATURE REVIEW

Vocabulary Knowledge and Lexical Sources of Difficulty in L2 Reading

Vocabulary knowledge is a robust correlate of L2 reading comprehension (Jeon & Yamashita, 2014; Nation, 2006; Qian, 1999, 2002; Schmitt et al., 2011). Tunmer and Chapman (2012), for example, treat vocabulary as part of oral language comprehension, while related work also suggests that vocabulary may contribute to reading comprehension beyond decoding alone. Studies of adolescent and young adult readers likewise show that vocabulary knowledge can strongly predict reading comprehension, suggesting that semantic knowledge plays an important role in building meaning from print (Braze et al., 2007).

Vocabulary knowledge has been conceptualized in various ways, but most frameworks treat it as a multidimensional construct rather than a unitary ability. Nation (2001), for example, defined word knowledge as involving form, meaning, and use, suggesting that knowing a word involves more than recognizing its form or recalling a single meaning. According to Chapelle (1998), vocabulary knowledge is defined in terms of four traits—vocabulary size, knowledge of word characteristics, lexicon organization, and processes of lexical access. Henriksen (1999) conceptualized it as consisting of

precision of knowledge, depth of knowledge, and receptive and productive knowledge. In Qian's (2002) framework, vocabulary knowledge includes four interrelated dimensions such as vocabulary size, depth, lexical organization, and the automaticity of receptive and productive lexical use. These dimensions interact during language use and development, and their relative importance may vary depending on the purpose of language use.

Beyond how vocabulary knowledge is conceptualized, its importance for reading lies in how it supports meaning access and integration during comprehension. Vocabulary knowledge supports efficient word recognition, access to contextually appropriate meanings, and the construction of a coherent text representation (Perfetti, 2007; Perfetti & Stafura, 2014). However, its contribution to comprehension depends on whether lexical knowledge is sufficiently precise and whether learners possess the resources needed to use it effectively. More specifically, comprehension may also break down when learners recognize a word form but cannot select or integrate the meaning required by the local and broader discourse context (Kintsch, 1988; Laufer, 1989a, 1990, 1997; Perfetti, 2007; Perfetti & Stafura, 2014). In this sense, word knowledge supports reading not only by enabling word identification but also by linking identified word meanings to sentence- and text-level comprehension (Nation, 2001, 2006; Perfetti & Stafura, 2014).

More broadly, threshold perspectives suggest that successful L2 reading depends on reaching a sufficient level of L2 language knowledge. Early work argued that reading problems in L2 may result from limited L2 proficiency rather than from a general lack of reading ability alone (Alderson, 1984; Clarke, 1979; Cummins, 1979). Consistent with this logic, contextual inferencing is also constrained by overall proficiency: even when less proficient learners report increased strategy use after instruction, their inferred meanings may remain inaccurate and fail to integrate with the broader text representation (Laufer, 1997; Shen & Wu, 2009). These patterns suggest that threshold accounts help explain when lexical resources become a bottleneck for higher-level comprehension processes. Thus, when learners' L2 proficiency is below a certain threshold, limited lexical knowledge may reduce learners' ability to infer word meanings and integrate those meanings into the developing text representation.

One influential way of specifying these threshold-related constraints is through lexical coverage, or the extent to which the words in a text are known. When learners do not know enough words in a text, local breakdowns may accumulate and interfere with global comprehension (Hu & Nation, 2000; Laufer, 1989b, 1992, 2020; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt et al., 2011). These studies converge on the view that readers below a certain level of lexical coverage often struggle with comprehension, whereas readers above that level are more likely to process texts effectively. From this view, greater vocabulary knowledge is advantageous because a higher density of known words makes contextual cues more usable and increases the likelihood that unfamiliar items can be interpreted successfully (Laufer, 1997; Webb, 2021). More broadly, lexical coverage has been shown to be closely associated with comprehension across contexts (Schmitt et al., 2011), although recommended coverage thresholds vary across studies, with differences often linked to text characteristics, modality (e.g., spoken vs. written), and the inferential demands required for adequate understanding (Hu & Nation, 2000; Laufer, 1989b, 1992, 2020; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt et al., 2011).

These findings suggest that lexical coverage explains an important source of comprehension difficulty in L2 reading, but it does not fully account for all forms of lexical difficulty. Lexical coverage does not operate mechanically: even when overall coverage is relatively high, comprehension support may depend more directly on the amount of attention devoted to individual unknown words than on coverage level itself (Pellicer-Sánchez et al., 2024). Likewise, lexical inferencing can still fail when the immediate context surrounding a target word contains multiple unknown items, because local context may provide insufficient cues for accurate meaning inference (Nassaji, 2003a). At the same time, lexical coverage has been found to be a consistent but weak predictor once vocabulary size and L2 reading comprehension were controlled, suggesting that other dimensions of word knowledge, such as vocabulary depth, may also contribute to comprehension (Song & Reynolds, 2022). Taken together, lexical coverage and threshold-based accounts primarily explain comprehension difficulty arising from insufficient L2 knowledge, whether at the level of word knowledge or broader language proficiency. However, they are less suited to explaining cases in which DT words trigger a plausible but context-inappropriate meaning (Laufer, 1989a, 1990, 1997). DT words may therefore be understood as a different kind of lexical challenge, one that involves not the absence of word knowledge per se, but imprecise or context-inappropriate meaning selection from a familiar form. This limitation makes it necessary to distinguish unknown-word difficulty from DT-related misinterpretation and provides the rationale for examining DT words as a potentially distinguishable lexical challenge in L2 reading.

Lexical Quality, Deceptively Transparent (DT) Words, and Meaning Construction

DT words become theoretically important once reading is viewed not simply as word recognition, but as an ongoing process of meaning selection and integration. In reading comprehension, learners must build a coherent mental representation of a

text by generating inferences, suppressing competing interpretations, and coordinating multiple sources of information under proficiency constraints (Kintsch, 1988; Nassaji, 2003b). From this perspective, comprehension may fail not only because a word is unknown, but also because a familiar form is assigned a contextually inappropriate meaning that is not adequately monitored or revised during reading (Kintsch, 1988; Laufer, 1989a, 1997; Perfetti, 2007; Perfetti & Stafura, 2014). Contextual clues do not guarantee accurate meaning construction for all learners, and reliance on context may still yield partial, unstable, or inaccurate lexical interpretations (Bensoussan & Laufer, 1984; Elgort, 2017; Frantzen, 2003; Haynes, 1993; Kelly, 1990). When a familiar-looking form activates a plausible but inappropriate meaning, the resulting local error may bias subsequent propositional integration and weaken text-level coherence. This problem shifts attention from lexical availability alone to the precision with which learners represent and use word meanings in context.

The Lexical Quality Hypothesis (LQH) offers a useful framework for considering how DT words may create comprehension difficulty even when learners experience a sense of familiarity. LQH proposes that successful reading depends on well-specified lexical representations in which orthographic, phonological, and semantic information is sufficiently precise and well integrated to support accurate access and flexible meaning construction during ongoing reading (Perfetti & Hart, 2002; Perfetti & Stafura, 2014). Additionally, Gor et al. (2022) suggest that lexical difficulty in L2 may arise not only from unknown words but also from imprecise lexical representations that weaken form-meaning connections and promote inappropriate meaning activation. From this perspective, the problem posed by DT words may interfere with text comprehension because learners may access a familiar form without establishing a sufficiently precise, context-appropriate meaning. In this sense, DT difficulty may be viewed as more closely related to lexical quality than to a simple case of lexical absence. This account also provides a rationale for examining context-sensitive meaning selection more directly.

Psycholinguistic and vocabulary research further suggests a plausible processing route through which DT words may lead to miscomprehension. Familiar lexical forms tend to activate dominant, salient, or otherwise readily available meanings early in processing (Duffy et al., 1988; Laufer, 1989a, 1997; Morton, 1969; Perfetti & Hart, 2002). Within the Construction-Integration model, such early candidates must be evaluated against the developing text representation, with contextually appropriate meanings maintained and less compatible alternatives suppressed (Kintsch, 1988). Related evidence further suggests that efficient word-level processing supports broader comprehension, whereas increased text difficulty can make word recognition and lexical access more effortful and thereby reduce overall understanding (Bell & Perfetti, 1994). Research on comprehension failure similarly indicates that word-level problems can cascade into broader failures of inference and coherence building (Cain & Oakhill, 1999, 2014; Perfetti et al., 1996). Relatedly, prior work on multiword meaning has shown that formally familiar items may diverge from the meanings of their constituent parts, so apparent transparency does not guarantee accurate interpretation (Moon, 1997; Wray, 2002). Research on lexical learning from context likewise suggests that learners' early lexical hypotheses are not always refined immediately through contextual exposure alone and that gains in semantic precision may remain uneven across encounters (Haynes, 1993; Pigada & Schmitt, 2006).

Research specifically on DT words is broadly consistent with this interpretation and suggests that deceptive familiarity can mislead learners during reading. Laufer (1989a, 1990, 1997) argued that DT words mislead learners when their surface forms, internal structure, or associated cues encourage an interpretation that appears plausible but does not match the contextually intended meaning. In this sense, the learner is not simply ignorant of the item; rather, the lexical representation available to the learner is incomplete, weakly specified, or misleading for the immediate context. Laufer (1997) further grouped DT words into several categories, including morphologically deceptive words, idioms, false cognates, polysemous words, and synforms. Although these categories differ in surface characteristics, they share a common risk: familiarity directs learners toward a meaning that is readily available but contextually inappropriate. Related research on polysemous words, form-similar words, morphological processing, and word knowledge also suggests that formal familiarity does not always guarantee accurate meaning selection (Abou-Khalil et al., 2019; Cieślicka, 2006; Logan & Kieffer, 2017; Nagy, 1995; Nagy et al., 1989; Nation, 2001; Rastle et al., 2004; Tyson, 1993; Zhang & Koda, 2012). Empirical findings reinforce this point. Laufer (1989a) found that DT words produced more comprehension errors than non-DT words, in part because learners were less likely to recognize DT items as problematic, and Laufer (1989a, 1990, 1997) further showed that misinterpretation of DT words was associated with poorer comprehension and translation performance. Laufer-Dvorkin (1985) reported similar results, while Bensoussan and Laufer (1984) found that EFL learners often relied on what they believed a word meant and turned to contextual guessing only when that initial interpretation broke down. Collectively, these findings suggest that DT difficulty may arise not because interpretation is blocked, but because interpretation proceeds on the basis of misleading familiarity.

Vocabulary Breadth, Vocabulary Depth, and Boundary Conditions in DT Text Comprehension

Vocabulary knowledge is widely understood as a multidimensional construct, and the distinction between vocabulary breadth and vocabulary depth is especially useful for clarifying the lexical demands of L2 reading comprehension (Qian, 1999, 2002; Schmitt, 2008). Different frameworks have conceptualized these components in different ways (Chapelle, 1998; Henriksen, 1999; Qian, 2002), but this breadth–depth distinction is particularly useful here because it helps specify what kind of lexical knowledge may matter most once words are encountered in context. Vocabulary breadth generally refers to how many words learners know, whereas vocabulary depth refers to how well those words are known in terms of semantic relations, multiple meanings, collocations, and other aspects of lexical organization (Anderson & Freebody, 1981; Nation, 2001; Nation & Beglar, 2007; Qian, 1999, 2002; Read, 1993, 2000, 2004a; Schmitt, 2008; Webb, 2012). Both dimensions are important for reading comprehension, and a growing body of evidence suggests that vocabulary breadth and depth are both associated with reading comprehension and share considerable overlap in the variance they explain (Alderson, 2000; Jeon & Yamashita, 2014; Laufer, 1992; Li & Kirby, 2015; Qian, 1999, 2002; Vermeer, 2001; Zhang & Yang, 2016). Yet they are not identical in function: breadth is more closely tied to lexical coverage (Laufer, 1992; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt et al., 2011), whereas depth is more closely tied to the precision, interconnectedness, and flexibility of lexical representations, which are needed for context-sensitive meaning selection and integration during reading (Perfetti, 2007; Perfetti & Hart, 2002; Qian, 2002; Read, 2004a). From this perspective, DT difficulty may not be fully captured by either vocabulary breadth or vocabulary depth alone, at least as these constructs are traditionally operationalized, because it involves not only lexical availability but also context-sensitive meaning selection from familiar-looking forms. Thus, the distinction between vocabulary breadth and vocabulary depth provides a basis for examining whether DT text reading comprehension is more closely associated with general lexical coverage, lexical precision, or both.

Vocabulary breadth may serve as an important baseline predictor of DT text comprehension because comprehension of connected text still depends on knowing a sufficient proportion of the surrounding vocabulary. A substantial body of research has shown that vocabulary breadth is a robust predictor of reading comprehension, largely because learners with larger vocabularies are more likely to know enough of the running words in a text to support local understanding and global integration (Dong et al., 2020; Jeon & Yamashita, 2014; Laufer, 1997; Laufer & Ravenhorst-Kalovski, 2010; Nurweni & Read, 1999; Qian, 1999, 2002; Schmitt et al., 2011; Stæhr, 2008; Tannenbaum et al., 2006). In this sense, breadth functions as a broad lexical resource that reduces the number of items readers must infer or bypass during reading. This role remains relevant in DT texts because DT words appear within connected passages whose overall comprehension still depends on general lexical coverage (Laufer, 1989b, 1992, 2020). If lexical coverage is too low, readers may struggle regardless of whether the main difficulty comes from unknown words or from deceptively familiar ones (Hu & Nation, 2000; Laufer, 1989b, 1992, 2020; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt et al., 2011). Breadth, then, is best treated as an important baseline condition in DT reading research, even though it may not by itself explain DT-related difficulty.

Vocabulary depth may be especially relevant once learners have gained enough access to the text, because DT comprehension depends on selecting meanings precisely rather than merely recognizing forms. This interpretation is consistent with research showing that vocabulary depth contributes uniquely to reading comprehension beyond breadth and word identification (Proctor et al., 2012). Depth has been associated with richer semantic knowledge, stronger connections among related words and meanings, and greater ability to distinguish among competing interpretations (Read, 1993, 1998, 2000, 2004a; Webb, 2012; Wolter, 2001; Zhang & Koda, 2017). These properties are directly relevant to DT processing because learners must do more than recognize a familiar form; they must determine which meaning fits the immediate context. From this perspective, breadth may help learners access the text, whereas depth may be more directly related to accurate interpretation once the text has been accessed (Y. Kang et al., 2012; Qian, 1999). This distinction may be especially relevant to DT items, which may require learners to select a context-appropriate meaning from a familiar-looking form. Accordingly, DT text comprehension may provide a context in which depth may be particularly consequential because the central problem is not simply word recognition, but context-appropriate sense selection.

The relative contribution of breadth and depth is therefore best understood as conditional on the interpretive demands of the task and on the learners' ability to monitor, revise, and integrate meaning during reading. Research has repeatedly shown that breadth and depth do not contribute identically across reading outcomes. Qian (1999, 2002) treated them as closely related but distinct aspects of vocabulary knowledge, and Ouellette (2006) similarly argued that vocabulary supports reading through more than one pathway, with meaning-related aspects being especially important for comprehension. Li and Kirby (2015) further found that breadth was more strongly associated with multiple-choice reading comprehension, whereas depth was more strongly associated with summary writing, which places heavier demands on deeper text processing. Zhang and Yang (2016) likewise reported that depth of vocabulary knowledge played an important role in L2 reading comprehension,

particularly when comprehension required more than surface-level recognition. These task-related differences are relevant to DT texts because learners may need to evaluate whether an initially available meaning is compatible with the surrounding context, a process that requires monitoring and possible revision.

When initially activated meanings are potentially misleading, successful comprehension may also depend on monitoring and repair processes. Background knowledge may sometimes compensate for limited L2 proficiency (Hudson, 1982), but such compensation is fragile when learners do not detect and revise lexical misunderstandings. Perfetti et al. (1996) emphasized that readers must monitor comprehension and repair misinterpretations as they read, and Perfetti (1986) likewise highlighted the importance of lexical access, working memory, and schemata in reading performance. Similarly, Cain and Oakhill (1999) showed that less skilled comprehenders often struggle with inferencing and integration. Research on lexical inferencing points in the same direction: successful inferencing depends on the coordinated use of multiple knowledge sources (Nassaji, 2003b), while vocabulary depth is closely related to learners' use of inferencing strategies and to inferencing success (Nassaji, 2006). Logan and Kieffer (2017) further found that knowledge of context-appropriate senses of common words uniquely predicted reading comprehension, while Abou-Khalil et al. (2019) showed that learners may have difficulty identifying the intended senses of familiar polysemous words. Taken together, these findings suggest that depth may become more important when comprehension depends on rejecting a plausible but inappropriate meaning and revising it in light of contextual evidence.

This issue is particularly important in Korean EFL contexts, where the general contribution of vocabulary to reading is well established but the lexical demands of DT texts remain underexplored. Korean studies have also reported close links between vocabulary knowledge and reading comprehension, including work that explicitly compared breadth and depth among Korean learners (H. Choi, 2013; Y. Kang et al., 2012). More recent research has also shown that vocabulary depth contributes to reading comprehension alongside other linguistic resources such as grammar (Y. Hur & Y. Kang, 2024). However, the available Korean evidence has largely addressed general reading comprehension rather than DT-related comprehension specifically. As a result, it remains unclear whether the relative importance of breadth and depth changes when Korean high school EFL learners read texts containing DT words, where the difficulty may involve not only unknown vocabulary but also the need to avoid misinterpretation of familiar-looking items. For that reason, comparing breadth and depth in DT text comprehension can help clarify whether DT-related difficulty is more closely related to general lexical coverage, depth-related lexical knowledge, or both.

Taken together, the gaps motivate the present study's examination of whether DT words affect connected-text comprehension, whether DT sentence-interpretation ability predicts DT text reading comprehension, and how vocabulary breadth and depth contribute to DT text reading comprehension.

The Present Study

The present study examines DT text reading comprehension among Korean high school EFL learners, focusing on the presence of DT words, DT sentence interpretation, and the relative contributions of vocabulary breadth and vocabulary depth. Using a within-subject, counterbalanced Latin square design that partially balanced task order and text topic across participants, it compares learners' text-level reading comprehension performance between a control text condition and a DT text condition. It also tests whether learners' performance on the DT sentence interpretation task (DT Sent) explains additional variance in DT text reading comprehension (DT RC) beyond relevant task- and learner-related factors, including Task Order, Text Topic, control text reading comprehension (CT RC), vocabulary breadth (VB), vocabulary depth (VD), and background knowledge (BK). Finally, the study assesses the relative contributions of VB and VD to DT RC after controlling for Task Order, Text Topic, CT RC, and BK. Through these analyses, the study examines whether and to what extent DT text reading performance is associated with DT-specific sentence interpretation and vocabulary knowledge after accounting for general reading ability, background knowledge, Task Order, and Text Topic. The following research questions guide the study:

1) Do Korean 11th-grade EFL learners' reading comprehension scores differ between the CT and DT conditions?
2) To what extent does performance on the DT Sent task explain additional variance in DT RC beyond Task Order, Text Topic, CT RC, VB, VD, and BK?
3) What are the relative contributions of VB and VD to DT RC after controlling for Task Order, Text Topic, CT RC, and BK?

METHOD

Participants

Data were collected from 40 Korean 11th-grade (second-year high school) students at an all-boys public high school in Jeonbuk Special Self-Governing Province, Republic of Korea. All participants reported Korean as their first language. Their mean age was 17.61 years (*SD* = 0.59). Participants' English proficiency was described using their scores on the September national mock exam. The mean exam score was 48.23 (*SD* = 21.29) out of 100 points, slightly below the national mean of 57.98 (Seoul Metropolitan Office of Education, 2025). Based on the exam's nine-band grading system, 15% of students were in the high band (Grades 1–3), 52.5% in the middle band (Grades 4–6), and 32.5% in the low band (Grades 7–9). Compared with the national distribution (high = 34.04%, middle = 41.59%, low = 23.37%; Seoul Metropolitan Office of Education, 2025), this sample included fewer high-band students and more low-band students. The proficiency profile should be considered when interpreting the reading comprehension outcomes.

Measures

Vocabulary Breadth

To assess participants' vocabulary breadth (VB), the Vocabulary Size Test (VST; Nation & Beglar, 2007) was used. The VST is a widely used standardized measure of receptive vocabulary breadth in L2 research and has demonstrated strong reliability in prior studies. Importantly, the VST is based on the word-family principle; thus, vocabulary breadth was operationalized as receptive knowledge of English word families. The VST samples vocabulary from the 14,000 most frequent English word families, organized into fourteen 1,000-word-family levels, each represented by 10 multiple-choice items. By the end of Korean high school, students are expected to have learned approximately 2,500–3,000 words (Ministry of Education, 2022). Because this curricular estimate is not based on word families, it was used only as a rough curricular reference point in selecting the upper range of the VST. To maintain consistency with the VST's unit of measurement and to align the breadth measure with the lexical range from which DT words were selected, the test ceiling was set at the 3,000-word-family level (30 items), which was considered appropriate for the participants' proficiency level. In each item, the target word appears in capital letters on the left, followed by an example sentence containing that word and four multiple-choice options. Participants were asked to choose the closest meaning from the four options. While the VST is often administered without strict time limits (Nation & Beglar, 2007), a 15-minute limit was adopted to maintain testing efficiency for this participant group. Each item was worth one point, yielding a total score range of 0 to 30. Accordingly, VST scores were interpreted as an estimate of receptive knowledge of word families rather than individual word forms (see Appendix C).

Vocabulary Depth

To assess participants' vocabulary depth (VD), the Word Associates Test (WAT; Read, 1993) was administered. The WAT is a widely used measure of receptive vocabulary depth. In the present study, the WAT assessed participants' knowledge of the semantic and collocational associations of target English adjectives. These associations reflect both meaning-based relations and usage-based co-occurrence patterns, which are key components of vocabulary depth. The present study adopted the adjective-target format of the WAT because adjective targets are well suited to assessing semantic relations and collocational patterns, including paradigmatic and syntagmatic aspects of vocabulary depth as operationalized in the WAT (Read, 1993, 1998). Following the rationale reviewed by Zhang and Koda (2017), the use of a single target-word class was intended to increase item homogeneity and promote consistency in semantic relationships and response patterns. In each WAT item, a target word appears above eight options presented as checkboxes. Each correct option falls into one of three association types—paradigmatic (e.g., *rapid/quick*), syntagmatic (e.g., *strong/argument*), or analytic (e.g., *rapid/speed*). The four correct options typically represent a combination of these association types. Following the procedures of Qian (2002) and Read (1998), participants were required to complete the test within 30 minutes. They selected four options per item and received one point for each correct selection, with no penalty for incorrect choices. Thus, the maximum score was four points per item, yielding a total score range from 0 to 160 across the 40 items (see Appendix D).

Background Knowledge Questionnaire

Participants' background knowledge (BK) was measured using a self-report questionnaire informed by McNeil (2011). Six questionnaire items were prepared for each of the two topics. Of these, three items per topic were directly related to the content of the corresponding passage and were included in the BK scores, yielding six scored items in total. The remaining items served as fillers and were not included in the total score. The six scored items used a three-option response format (True, False, or I do not know) and were scored as 2 for a correct response, 1 for *I do not know*, and 0 for an incorrect response to reduce guessing bias. Thus, the total possible score ranged from 0 to 12. The scored items showed acceptable internal consistency (Cronbach's $\alpha = .73$; see Appendix B).

Reading Comprehension Test

To assess participants' reading comprehension, two base reading passages were adapted from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) 8th Edition Maze Benchmark, Grade 6 materials (University of Oregon, 2023). Each base passage was adapted into two text versions—a control text and a DT text—yielding four passages (CT1, DT1, CT2, and DT2). Although the original Maze test measures silent fluency and comprehension using a multiple-choice cloze task format, in the present study, each selected passage was modified from the original cloze format into 10 multiple-choice comprehension questions: seven factual, two inferential, and one requiring synthesis (i.e., integrating information across the passage to explain an event), following Alderson's (2000) categorization of question types.

The base texts were first prepared in a transparent form without DT words; these served as the control text versions. In this study, DT words and expressions were operationalized as items whose surface forms are familiar but whose context-appropriate meanings diverge from the meanings learners are likely to activate, thereby increasing the likelihood of misinterpretation. DT items were selected if they satisfied at least two of the following four criteria: (a) familiar surface form in terms of morphology and frequency, particularly within the top 3,000 English word families, (b) contextual sense diverges from the first dictionary meaning or from a straightforward compositional reading (e.g., polysemy, metaphorical extension, domain-specific/technical sense), (c) interpretation hinges on collocational conventions (e.g., fixed expressions, idioms, phrasal verbs), and (d) a functional shift or change in grammatical role within the sentence is likely to induce misinterpretation. Consequently, within each topic pair (CT1 and DT1, and CT2 and DT2), the intended distinction between the paired control and DT texts was the presence or absence of DT items. Across the two DT texts, DT coverage differed descriptively (DT1: 1.58%; DT2: 2.80%). DT coverage in the two DT texts was kept within approximately 1.5%–3.0% so that the DT manipulation would remain limited in scope while still allowing localized DT-related difficulty to be examined during reading comprehension. In the DT1 text, seven DT words (i.e., *leaf, block, even, about, stand out in sharp relief, place, apply the ink*) were embedded, accounting for 1.58% of 449 total words. In the DT2 text, 15 DT words and expressions were inserted (2.80% of 536 total words; *barn, free, last, plain, term, hit, dump, drift, story, severe, strike, ground, stock up, take lightly, historic*). Because DT coverage differed across the two DT texts, Text Topic (i.e., the topic of the DT text read: DT1 vs. DT2) was included as a control variable in the regression analyses. DT coverage was calculated as the proportion of DT word tokens relative to total running words in each text. The readability of the two DT texts was comparable, with Flesch Reading Ease scores of 71.6 and 70.4, respectively. In addition, the four texts showed similar readability levels (CT1 = 71.7, DT1 = 71.6, CT2 = 69.6, and DT2 = 70.4).

To reduce test fatigue, the reading tests were administered across two sessions; the first session included one reading test, and the second session was conducted six days later. Each participant read one text per session and answered ten multiple-choice comprehension items within a 20-minute time limit. Each item included one correct answer and four distractors and was worth one point, yielding a possible total score of 0 to 10. Scores on the control text and DT text reading comprehension tests were used as CT RC and DT RC, respectively. Two counterbalanced test forms were created: Form A (CT1 + DT2) and Form B (CT2 + DT1). Internal consistency reliability was computed for each form (20 items). Form A showed acceptable reliability (KR-20 = .72) and Form B demonstrated good reliability (KR-20 = .82; see Appendix E).

DT Sentence Interpretation Task

For the DT sentence interpretation task, participants completed five sentence-embedded DT items drawn from the DT text they had just read. The task was administered immediately after the DT text reading comprehension test. In each item, one DT word was underlined. For each underlined word, participants first rated their familiarity on a three-point scale (0 = unknown, 1 = acquainted, 2 = known) and were then asked to translate its meaning into Korean based on the sentence context. Participants were given 15 minutes to complete the task. This two-step procedure was adapted from Laufer (1989a, 2020) to elicit both self-perceived word

knowledge and context-based meaning interpretation. The self-rating responses were collected as supplementary information, but in the present analyses, DT Sent was operationalized using only the translation scores. To reduce participant burden, not all DT words in the two texts were tested. Instead, five DT target items were selected from each DT text, so that both DT texts were represented in the overall design. Selection followed three criteria. First, the selected items were restricted to words within the 3,000 most frequent English word families (Laufer, 1992). Second, they were classified as DT items according to Laufer's (1997) definition of DT words. Third, words were selected for their contextual importance to overall text understanding (Laufer & Ravenhorst-Kalovski, 2010). Accordingly, DT Sent was treated as a targeted measure of how learners interpreted selected DT items in context, rather than as a comprehensive measure of all DT-related processing demands in the texts. Responses were scored on a three-point rubric, 0 for incorrect, 1 for partially correct, and 2 for fully correct. For each participant, DT Sent was calculated as the total translation score across the five DT items from the DT text read by that participant, yielding a score range of 0 to 10. Higher scores indicated more accurate contextual interpretation. Two independent raters evaluated the translations, and disagreements were resolved through discussion. Inter-rater agreement was assessed with Cohen's κ. To estimate inter-rater reliability, a random 20% subsample of the data (n = 8; 20% of $N = 40$) was double-rated, and inter-rater reliability on this subsample was high (κ = .85; see Appendix F).

Procedures

All measures were administered in a group setting. Prior to testing, participants received standardized instructions and completed brief practice items to familiarize themselves with each task format. Data collection proceeded in the following order: (a) a vocabulary breadth measure, (b) a vocabulary depth measure, (c) a background knowledge questionnaire, (d) two multiple-choice reading comprehension tests (one for the control text and one for the DT text), and (e) a DT sentence interpretation task. To reduce fatigue, the two reading comprehension tests were administered across two sessions, with the second session conducted six days after the first. In each session, participants read one text and completed the corresponding 10-item comprehension test within a 20-minute time limit. In the session in which participants read the DT text (DT1 or DT2), they completed the DT sentence interpretation task immediately after the DT text reading comprehension test.

The study employed a within-subject, counterbalanced Latin square design in which the CT text condition and DT text condition were partially balanced across Task Order (DT-first vs. CT-first) and Text Topic. Specifically, two topic-specific text pairs were prepared (CT1-DT1 for Topic 1; CT2-DT2 for Topic 2), with DT words appearing only in the DT versions. Each participant read one CT text and one DT text from different topics (i.e., CT1 + DT2 or CT2 + DT1). Task Order was counterbalanced across participants. Given this partial balancing, Task Order and Text Topic (i.e., the topic of the DT text read: DT1 vs. DT2) were included as control variables in the regression analyses.

Data Analysis

The data were analyzed using IBM SPSS Statistics (Version 29.0.2). First, descriptive statistics were computed for all study variables, including CT RC, DT RC, DT Sent, VB, VD, and BK. For each measure, minimum and maximum scores, means, and standard deviations were calculated. Pearson correlations were then calculated to examine bivariate associations among the variables. Because the study used a within-subject design, each participant completed both the control text condition and the DT text condition. To address RQ1, a paired-samples *t*-test was conducted to compare mean reading comprehension scores between CT RC and DT RC. To address RQ2, a hierarchical multiple regression analysis was conducted with DT RC as the dependent variable. Model 1 included Task Order (DT-first vs. CT-first), Text Topic (DT1 vs. DT2), CT RC, VB, VD, and BK, and Model 2 added DT Sent. To address RQ3, a second hierarchical regression analysis was conducted with DT RC as the dependent variable. Model 1 included Task Order, Text Topic (i.e., the topic of the DT text read: DT1 vs. DT2), CT RC, and BK. Model 2 added VB, and Model 3 added VD. In both hierarchical regression analyses, Task Order and Text Topic were entered as control variables rather than focal predictors. Tolerance and variance inflation factor (VIF) were examined in the regression models to assess multicollinearity. All analyses were two-tailed, with the alpha level set at .05.

FINDINGS AND DISCUSSION

Descriptive Statistics of Study Variables

Table 1 summarizes participants' performance on the six target study measures. On the two reading comprehension measures, participants scored an average of 4.00 out of 10 ($SD = 2.67$) on CT RC and 3.88 out of 10 ($SD = 2.88$) on DT RC. These results

show that the mean scores were similarly low across the two reading conditions. On the vocabulary measures, VB scores ranged from 3 to 26 out of 30 (M = 14.12, SD = 6.57), whereas VD scores ranged from 56 to 107 out of 160 (M = 77.30, SD = 10.27). On DT Sent, the participants obtained a mean score of 3.10 out of 10 (SD = 3.18), indicating generally low performance together with substantial variation across individuals. This low mean score should be considered when interpreting subsequent analyses involving DT Sent. By contrast, BK scores ranged from 0 to 12, with a relatively high mean of 8.38 (SD = 2.57), suggesting that the participants obtained relatively high scores on the topic-related background knowledge measure.

TABLE 1

Descriptive Statistics of Study Variables (N = 40)

Variable	Min.	Max.	M	*SD*
CT RC	0	10	4.00	2.67
DT RC	0	10	3.88	2.88
VB	3	26	14.12	6.57
VD	56	107	77.30	10.27
DT Sent	0	10	3.10	3.18
BK	0	12	8.38	2.57

Note. CT RC = control text reading comprehension; DT RC = DT text reading comprehension; VB = vocabulary breadth; VD = vocabulary depth; DT Sent = DT sentence interpretation task; BK = background knowledge.

The Relationships among Study Variables

To examine associations among the tested measures, Pearson correlation coefficients were calculated for all study variables (see Table 2). The two control variables, Task Order and Text Topic, were not significantly associated with any of the main measures; for brevity, these coefficients are not reported in Table 2. Among the main measures, CT RC and DT RC showed a moderate positive correlation (r = .57, p < .001), indicating that participants who performed better on one reading comprehension measure also tended to perform better on the other. VB and VD were both significantly correlated with CT RC (r = .50, p < .001; r = .45, p < .01) as well as with DT RC (r = .54, p < .001; r = .73, p < .001). VD showed the strongest association with DT RC; at the bivariate level, this pattern suggests that VD may be more closely related to DT RC than VB. VB and VD were also strongly correlated with each other (r = .76, p < .001), representing the highest correlation among the measured variables. This pattern suggests substantial shared variance between the two vocabulary constructs, although they remained analytically distinct in the study.

By contrast, DT Sent showed no significant correlations with the other main measures (rs = .21–.25), indicating that DT Sent was not significantly associated with the broader reading and vocabulary measures. In addition, BK was significantly correlated only with CT RC (r = .34, p < .05) and was not significantly related to DT RC, VB, VD, or DT Sent. Taken together, these bivariate results suggest that lexical knowledge, particularly vocabulary depth, was more consistently related to reading comprehension performance than performance on the DT sentence interpretation task or background knowledge in the present sample.

TABLE 2

Pearson Correlations among Study Variables (N = 40)

Variable	1	2	3	4	5
1. CT RC	1				
2. DT RC	.57***	1			
3. VB	.50***	.54***	1		
4. VD	.45**	.73***	.76***	1	
5. DT Sent	.25	.21	.24	.23	1
6. BK	.34*	.21	.15	.03	.24

Note 1. CT RC = control text reading comprehension; DT RC = DT text reading comprehension; VB = vocabulary breadth; VD = vocabulary depth; DT Sent = DT sentence interpretation task; BK = background knowledge.

Note 2. *p < .05. **p < .01. ***p < .001.

RQ1: Mean Differences in Reading Comprehension Between CT and DT Conditions

To address RQ1, a paired-samples *t*-test was conducted to compare participants' scores on CT RC and DT RC. In the within-subject comparison, participants scored a mean of 4.00 out of 10 (*SD* = 2.67) on CT RC and 3.88 out of 10 (*SD* = 2.88) on DT RC. The observed mean difference between the two text conditions was therefore small (.12 points). This difference was not statistically significant, $t(39) = .31$, $p = .76$, indicating that the DT condition did not produce a statistically significant mean decrease in text-level reading comprehension compared with the CT condition.

The non-significant result for RQ1 should not be interpreted simply as evidence that DT words had no effect on reading comprehension. Rather, the non-significant CT–DT difference should be interpreted in light of the boundary conditions of the present design. Participants' low mean scores in both CT RC and DT RC suggest that many learners may have experienced general comprehension difficulty across the two conditions. This interpretation is consistent with the sample's proficiency profile: participants' mean score on the September mock exam was below the national mean, and the sample was concentrated in the middle-to-lower proficiency bands. In this learner group, limited resources for constructing text-level comprehension may have made the additional difficulty associated with DT words less likely to emerge as a statistically detectable condition effect. Therefore, the present finding should not be generalized as evidence that DT words do not affect L2 reading comprehension across learner populations.

This proficiency-related explanation is also compatible with proficiency-threshold and lexical coverage perspectives. From these perspectives, learners may need sufficient linguistic and lexical resources to construct a stable text-level representation before more specific sources of comprehension difficulty become observable (Alderson, 1984; Clarke, 1979; Hu & Nation, 2000; Laufer, 1989b, 2020; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006). Importantly, because text-specific lexical coverage was not directly calculated in the present study, this explanation should be treated as tentative. When overall comprehension is already low, as in the present study, broader comprehension difficulty may overshadow the additional processing burden associated with DT words. Accordingly, future research should use proficiency-stratified designs to clarify whether and under what proficiency conditions CT–DT differences in reading comprehension emerge.

TABLE 3

Descriptive Statistics and Paired-Samples T-test Results for CT RC and DT RC (N = 40)

RC Measures	Min.	Max.	M	*SD*	*t*	*p*
CT RC	0	10	4.00	2.67	.31	.76
DT RC	0	10	3.88	2.88		

Note. RC = reading comprehension; CT RC = control text reading comprehension; DT RC = DT text reading comprehension.

RQ2: Predicting DT Text Reading Comprehension from Performance on the DT Sentence Interpretation Task

To address RQ2, hierarchical regression analyses were conducted to examine whether performance on the DT Sent task, which indexed sentence-level meaning selection, explained additional variance in DT RC, which indexed text-level comprehension, after the effects of other relevant predictors had been taken into account (see Table 4). In Model 1, Task Order, Text Topic, CT RC, VB, VD, and BK were entered as predictors. The model was statistically significant, $R^2 = .63$, adjusted $R^2 = .56$, $F(6, 33) = 9.33$, $p < .001$, indicating that these predictors together accounted for 63% of the variance in DT RC. Among the predictors, CT RC and VD emerged as significant positive predictors, whereas VB, BK, Task Order, and Text Topic were not significant.

In Model 2, the analysis then tested whether DT Sent contributed uniquely beyond the predictors already included in Model 1. After DT Sent was added, the overall model remained significant, $R^2 = .63$, adjusted $R^2 = .55$, $F(7, 32) = 7.77$, $p < .001$. However, the addition of DT Sent did not yield a significant increase in explained variance ($\Delta R^2 = .00$, $\Delta F(1, 32) = .04$, $p = .85$). DT Sent did not account for additional variance in DT RC ($B = -.02$, $\beta = -.02$, $p = .85$). In this second model, VD remained significant, and CT RC was at the conventional significance threshold ($p = .050$). This pattern was also consistent with the bivariate analysis, in which DT Sent was not significantly correlated with DT RC. These results indicate that DT Sent did not explain unique variance in DT RC under the present model and measurement conditions.

One plausible interpretation of this pattern is that the two measures captured different levels of comprehension. The DT Sent task primarily assessed local, sentence-level meaning selection, whereas DT RC required readers to construct and maintain a coherent representation across the text. This pattern may reflect a mismatch between the level targeted by DT

Sent and the broader integration demands of DT RC. Performance on DT Sent may therefore not translate directly into successful DT RC, because connected-text comprehension requires readers to integrate lexical and contextual information across sentences, monitor coherence, and revise earlier interpretations when necessary. In other words, successful DT text reading comprehension depends not only on accurate sentence-level interpretation but also on the ability to integrate propositions across sentences and maintain coherence over time (Cain & Oakhill, 1999; Kintsch, 1988; Landi & Ryherd, 2017; Nassaji, 2003b, 2006). Accordingly, similarities in sentence-level performance do not necessarily imply comparable text-level comprehension outcomes.

TABLE 4

Hierarchical Regression Analyses Predicting DT Text Reading Comprehension from Performance on the DT Sentence Interpretation Task (N = 40)

Model		Model 1			Model 2	
Regression Component	*B*	*β*	*t*	*B*	*β*	*t*
Task Order	.37	.07	.58	.38	.07	.58
Text Topic	.05	.01	.07	.07	.01	.11
CT RC	.30	.28	2.06*	.30	.28	2.03
VB	-.09	-.20	-1.08	-.09	-.19	-1.04
VD	.21	.74	4.33***	.21	.74	4.27***
BK	.15	.13	1.13	.15	.14	1.13
DT Sent	—	—	—	-.02	-.02	-.19
R^2		.63			.63	
Adj. R^2		.56			.55	
ΔR^2		—			.00	
F		9.33***			7.77***	
ΔF		—			.04	

Note 1. The dependent variable was DT RC. CT RC = control text reading comprehension; DT RC = DT text reading comprehension; VB = vocabulary breadth; VD = vocabulary depth; BK = background knowledge; DT Sent = DT sentence interpretation task.

Note 2. *$p < .05$. **$p < .01$. ***$p < .001$.

Note 3. Em dashes (—) indicate either predictors not entered in the corresponding model or model-comparison statistics that were not applicable because there was no preceding model for comparison.

This sentence-text distinction is particularly important for interpreting DT-related processing difficulty. When learners encounter a familiar form that activates a context-inappropriate meaning, the difficulty may influence the developing text representation unless it is detected and revised (Bensoussan & Laufer, 1984; Kintsch, 1988; Laufer, 1989a, 1990, 1997). Under this view, DT-related difficulty may not simply reflect an isolated sentence-level interpretation error. Rather, it may involve difficulty in maintaining coherence and updating interpretations across propositions, especially when learners' lexical representations are underspecified or when dominant but inappropriate meanings are strongly activated (Laufer, 1989a, 1997; Perfetti, 2007; Perfetti & Hart, 2002; Perfetti & Stafura, 2014).

A second explanation concerns the measurement characteristics of the DT Sent task, particularly sensitivity and construct coverage. In the present study, the task was based on only five sentence-embedded DT items drawn from the DT text read by each participant and was scored on a 10-point scale using translation responses. Participants' mean performance was also low (M = 3.10, *SD* = 3.18), which may have limited the task's ability to differentiate among participants. Accordingly, the lack of a significant incremental contribution from DT Sent may reflect both limited construct coverage and insufficient measurement sensitivity. In addition, although the task was intended to capture contextually appropriate interpretation of DT items, it may not have represented the full range of abilities involved in DT-related reading comprehension. More specifically, the task focused on local sentence-level interpretation of a limited set of selected DT items rather than the broader integration and monitoring processes involved in DT text reading comprehension. Thus, DT Sent should be interpreted as a targeted, translation-based measure of selected DT-item interpretation rather than as a comprehension measure of DT processing during connected-text reading. For these reasons, this result should be interpreted cautiously, because it may reflect not only a limited overlap between sentence-level DT interpretation and DT RC, but also limited construct coverage in the present measure.

Although DT Sent did not explain additional variance in DT RC, vocabulary depth remained a robust predictor of DT RC across the regression models. This finding is consistent with the possibility that more precise lexical knowledge is important for DT text reading comprehension, although lexical quality was not directly measured in the present study. This pattern motivates RQ3, which examines the relative contributions of VB and VD to DT RC.

RQ3: Relative Contributions of Vocabulary Breadth and Depth to DT Text Reading Comprehension

To address RQ3, hierarchical multiple regression analyses were conducted to determine the extent to which VB and VD contributed to explained variance in DT RC (see Table 5). In Model 1, Task Order, Text Topic, CT RC, and BK were entered as control variables. The model was statistically significant, $R^2 = .33$, $F(4, 35) = 4.38$, p = .006, indicating that the control variables together explained 33% of the variance in DT RC. Within this model, CT RC was the only significant predictor ($\beta = .54$, $p = .001$).

In Model 2, VB was added after the control variables. With the inclusion of VB, the proportion of explained variance increased from .33 to .42, and this increase was statistically significant, $\Delta R^2 = .09$, $\Delta F(1, 34) = 4.97$, $p = .033$. The overall model also remained significant, $F(5, 34) = 4.90$, $p < .01$. In this model, VB was a significant predictor ($\beta = .36$, $p = .033$), and CT RC also remained significant ($\beta = .37$, $p = .034$). At this stage, vocabulary breadth accounted for additional variance in DT RC beyond the control variables.

In Model 3, VD was entered after VB. This final model explained 63% of the variance in DT RC, $R^2 = .63$, $F(6, 33) = 9.33$, $p < .001$, and the increase in explained variance from Model 2 to Model 3 was statistically significant ($\Delta R^2 = .21$, $\Delta F(1, 33) = 18.73$, $p < .001$). In the final model, VD was the strongest predictor of DT RC ($\beta = .74$, $p < .001$), and CT RC remained a smaller but still significant positive predictor ($\beta = .28$, $p = .047$). By contrast, VB was no longer significant after VD was entered ($\beta = -.20$, $p = .288$). Notably, VB had been significant when entered before VD, but its coefficient became nonsignificant and reversed direction in the final model. The diagnostic statistics were within acceptable ranges, with VIF values ranging from 1.02 to 2.97 and tolerance values ranging from .34 to .98. The reversal in the VB coefficient should therefore be interpreted cautiously. Rather than indicating a true negative effect of vocabulary breadth on DT text reading comprehension, it may reflect the substantial shared variance between VB and VD. This interpretation is consistent with the strong bivariate correlation between VB and VD reported in Table 2.

Taken together, these findings indicate that the variance in DT RC initially captured by VB was largely shared with VD and was more clearly accounted for by vocabulary depth once both vocabulary variables were considered simultaneously. Under the present task conditions, VD accounted for a larger unique share of the variance in DT RC. This pattern is consistent with previous research showing that vocabulary depth explains unique variance in L2 reading comprehension and, in some contexts, may be a stronger predictor than vocabulary breadth (Y. Hur & Y. Kang, 2024; Y. Kang et al., 2012; Ouellette, 2006; Qian, 1999, 2002; Zhang & Yang, 2016). It is also consistent with accounts of vocabulary depth as a multidimensional construct involving more than basic form–meaning knowledge, including richer meaning representations, syntactic functioning, collocational possibilities, and register-related knowledge (Read, 2004a, 2004b; Wolter, 2001).

One plausible explanation is that DT texts in the present study may have placed relatively greater demands on precise sense selection than on vocabulary size alone. Because DT items are familiar in form but potentially misleading in meaning, learners may need to distinguish among related senses, suppress inappropriate interpretations, and maintain contextually appropriate meanings during ongoing reading (Laufer, 1989a, 1990, 1997). Thus, learners with richer lexical representations may have been better able to sustain these context-appropriate meanings and integrate them coherently across the text (Kintsch, 1988; Perfetti & Stafura, 2014; Qian, 2002; Read, 1993, 1998). This may help explain why VD, rather than VB, was the stronger predictor of DT RC in the final model. From this perspective, vocabulary depth may have reflected lexical resources involving semantic precision, relational knowledge, and context-appropriate meaning selection under the present task conditions (Laufer, 1997; Qian, 2002; Read, 1993, 1998).

At the same time, the present findings should not be taken to mean that vocabulary breadth is unimportant. Vocabulary breadth has been shown to explain substantial variance in reading comprehension across a range of learner groups and task types (Qian, 1999, 2002; Tannenbaum et al., 2006; Zhang & Yang, 2016). However, prior studies suggest that the relative contribution of breadth and depth depends on task demands. Breadth has often been linked more closely to tasks that rely on general lexical coverage, rapid recognition, or relatively literal comprehension, whereas depth has tended to become more predictive when successful performance depends on semantic differentiation, inferencing, and integration (Li & Kirby, 2015; Nassaji, 2006; Ouellette, 2006; Zhang & Yang, 2016). Thus, the present RQ3 findings are best interpreted as evidence that VD had a stronger unique contribution than VB under the present DT text conditions, not as evidence that vocabulary breadth is generally unimportant.

TABLE 5

Hierarchical Regression Analyses Predicting DT Text Reading Comprehension from Vocabulary Breadth and Depth (N = 40)

Model	Model 1			Model 2			Model 3		
Regression Component	*B*	*β*	*t*	*B*	*β*	*t*	*B*	*β*	*t*
Task Order	.49	.09	.60	.20	.04	.26	.37	.07	.58
Text Topic	-.08	-.01	-.10	.41	.07	.52	.05	.01	.07
CT RC	.59	.54	3.56**	.39	.37	2.22*	.30	.28	2.06*
BK	.04	.04	.26	.04	.03	.22	.15	.13	1.13
VB	—	—	—	.16	.36	2.23*	-.09	-.20	-1.08
VD	—	—	—	—	—	—	.21	.74	4.33***
R^2		.33			.42			.63	
Adj. R^2		.26			.33			.56	
ΔR^2		—			.09			.21	
F		4.38**			4.90**			9.33***	
ΔF		—			4.97*			18.73***	

Note 1. The dependent variable was DT RC. CT RC = control text reading comprehension; DT RC = DT text reading comprehension; BK = background knowledge; VB = vocabulary breadth; VD = vocabulary depth.

Note 2. *$p < .05$. **$p < .01$. ***$p < .001$.

Note 3. Em dashes (—) indicate either predictors not entered in the corresponding model or model-comparison statistics that were not applicable because there was no preceding model for comparison.

CONCLUSION

This study investigated Korean high school EFL learners' reading comprehension of texts containing deceptively transparent (DT) words. More specifically, it tested (1) whether CT RC differed from DT RC, (2) the extent to which learners' performance on the DT sentence interpretation task explained additional variance in DT RC, and (3) the relative contributions of vocabulary breadth and vocabulary depth to DT RC. Overall, three main findings emerged. First, CT RC and DT RC did not differ significantly under the present text conditions. Second, DT sentence interpretation performance did not explain additional variance in DT RC. Third, vocabulary depth was a stronger predictor of individual differences in DT RC than vocabulary breadth under the present task conditions.

The first finding was that CT RC and DT RC did not differ significantly. This result should not be interpreted as evidence that DT words are irrelevant to L2 reading comprehension. Rather, the lack of difference between the CT and DT conditions may reflect the present sample and task conditions, particularly the learners' low overall reading comprehension performance and middle-to-lower proficiency profile. Under these conditions, DT-related effects may have been difficult to detect as a statistically significant difference between conditions.

The second finding was that performance on the DT sentence interpretation task did not account for additional variance in DT RC once other predictors were taken into account. This result suggests that accurate interpretation of selected DT items at the sentence level did not translate into independent prediction of text-level DT reading comprehension. DT RC may require more than local meaning selection; it may also require learners to integrate meanings across sentences, monitor coherence, and revise interpretations during connected-text reading. At the same time, the lack of a significant incremental contribution from DT Sent should be interpreted cautiously because the task included only five selected DT items and relied on translation-based scoring. Thus, the result may reflect not only a limited overlap between sentence-level interpretation and text-level comprehension, but also the task's limited construct coverage and reduced sensitivity to individual differences.

The third finding was that vocabulary depth (VD) explained substantial unique variance in DT RC beyond vocabulary breadth (VB). VB initially accounted for additional variance, but its contribution was no longer significant once VD was entered. This pattern indicates that part of the variance initially associated with VB was shared with VD and that VD became the stronger predictor once both vocabulary measures were considered together. This result does not mean that vocabulary breadth is unimportant. Rather, it suggests that DT RC was more closely related to the kind of lexical knowledge indexed by vocabulary depth, such as semantic precision and relational knowledge, than to vocabulary breadth alone. Because the

present study did not directly measure lexical quality, however, this interpretation remains tentative and should be examined more directly in future research.

Although the present study did not directly examine instructional interventions, the findings may offer tentative pedagogical implications for Korean high school EFL contexts. Pedagogically, instruction addressing DT-related difficulty could provide opportunities for learners to examine context-dependent meanings of familiar-looking words. In exam-oriented reading classrooms, such instruction may include activities that require learners to compare multiple meanings, identify context-appropriate senses, and explain why an initially plausible meaning does not fit the surrounding text. These activities could help learners develop more precise lexical knowledge and more careful monitoring of meaning during reading. At the same time, vocabulary breadth remains important as a baseline resource for accessing texts. Therefore, vocabulary instruction may be better directed not toward replacing breadth-oriented vocabulary learning with depth-oriented work, but toward connecting the two by helping learners use known words more precisely in context.

Several limitations should be considered when interpreting the findings. One limitation concerns the study design. Although the study used a within-subject, counterbalanced Latin square design, Task Order and Text Topic were partially balanced rather than fully controlled. As a result, some order- and topic-related variation may have remained. Another limitation concerns the participant sample. The data were collected from learners in a single region, and the sample was not stratified by proficiency level. These features may limit the generalizability of the findings and may also have reduced the study's sensitivity to DT-related effects. Because the sample was concentrated in the middle-to-lower proficiency range, the non-significant CT–DT difference may reflect the difficulty of detecting DT-related effects under conditions of low overall reading performance, rather than showing that DT words have no general effect on L2 reading comprehension.

A further measurement limitation concerns the DT sentence interpretation task, which included only five selected DT items and used translation-based scoring. This may have limited the task's ability to capture the broader processes involved in DT text reading comprehension. Future research should therefore use designs that are more fully counterbalanced and stratified by proficiency, collect samples from more varied educational contexts, and use more comprehensive measures of DT interpretation to clarify whether and under what proficiency conditions DT effects are more or less likely to emerge.

On a final note, the present study contributes to L2 reading research by suggesting that, under the present conditions, DT-related difficulty should not be understood simply as a matter of whether DT words are present in a text. Rather, DT text reading comprehension should be interpreted in relation to learners' general comprehension level, lexical resources, and the measurement conditions under which DT difficulty is assessed. Taken together, these findings provide a cautious basis for future research on DT-related lexical difficulty in EFL reading contexts.

References

Abou-Khalil, V., Helou, S., Flanagan, B., Chen, M.-R. A., & Ogata, H. (2019). Learning isolated polysemous words: Identifying the intended meaning of language learners in informal ubiquitous language learning environments. *Smart Learning Environments*, *6*, Article 13. https://doi.org/10.1186/s40561-019-0095-0

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–27). Longman.

Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732935

Alderson, J. C., & Urquhart, A. H. (Eds.). (1984). *Reading in a foreign language*. Longman.

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). International Reading Association.

Bell, L. C., & Perfetti, C. A. (1994). Reading skill: Some adult comparisons. *Journal of Educational Psychology*, *86*(2), 244–255. https://doi.org/10.1037/0022-0663.86.2.244

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, *7*(1), 15–32. https://doi.org/10.1111/j.1467-9817.1984.tb00252.x

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, *40*(3), 226–243. https://doi.org/10.1177/00222194070400030401

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, *11*(5–6), 489–503. https://doi.org/10.1023/A:1008084120205

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique*, *114*(4), 647–662. https://doi.org/10.4074/S0003503314004035

Chapelle, C. A. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 32–70). Cambridge University Press. https://doi.org/10.1017/CBO9781139524711.004

Choi, Hyeyeon. (2013). Effects of depth and breadth of vocabulary knowledge on English reading comprehension among Korean high

school students. *Language Research*, *49*(2), 419–452.
Choi, Innchull. (2008). The impact of EFL testing on EFL education in Korea. *Language Testing*, *25*(1), 39–62. https://doi.org/10.1177/0265532207083744
Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, *22*(2), 115–144. https://doi.org/10.1191/0267658306sr263oa
Clarke, M. A. (1979). Reading in Spanish and English: Evidence from adult ESL students. *Language Learning*, *29*(1), 121–150. https://doi.org/10.1111/j.1467-1770.1979.tb01055.x
Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, *19*, 121–129.
Dong, Y., Tang, Y., Chow, B. W. Y., Wang, W., & Dong, W. Y. (2020). Contribution of vocabulary knowledge to reading comprehension among Chinese students: A meta-analysis. *Frontiers in Psychology*, *11*, Article 525369. https://doi.org/10.3389/fpsyg.2020.525369
Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*(4), 429–446. https://doi.org/10.1016/0749-596X(88)90066-6
Elgort, I. (2017). Incorrect inferences and contextual word learning in English as a second language. *Journal of the European Second Language Association*, *1*(1), 1–11. https://doi.org/10.22599/jesla.3
Frantzen, D. (2003). Factors affecting how second language Spanish students derive meaning from context. *The Modern Language Journal*, *87*(2), 168–199. https://doi.org/10.1111/1540-4781.00185
Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2022). Editorial: Fuzzy lexical representations in the nonnative mental lexicon. *Frontiers in Communication*, *7*, Article 1027692. https://doi.org/10.3389/fcomm.2022.1027692
Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd ed.). Routledge. https://doi.org/10.4324/9781315833743
Haynes, M. (1993). Patterns and perils of guessing in second language reading. In T. Huckin, M. Haynes, & J. Coady (Eds.), *Second language reading and vocabulary learning* (pp. 46–64). Ablex Publishing Corporation.
Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition*, *21*(2), 303–317. https://doi.org/10.1017/S0272263199002089
Hu, M. H.-C., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430. https://doi.org/10.64152/10125/66973
Hudson, T. (1982). The effects of induced schemata on the short circuit in L2 reading: Non-decoding factors in L2 reading performance. *Language Learning*, *32*(1), 1–33. https://doi.org/10.1111/j.1467-1770.1982.tb00516.x
Hur, Yuncheong, & Kang, Yusun. (2024). Relative contributions of vocabulary and grammar to Korean EFL learners' reading comprehension. *Foreign Languages Education, 31*(4), 27–50. https://doi.org/10.15334/FLE.2024.31.4.27
Jeon, E. Y., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. https://doi.org/10.1111/lang.12034
Kang, Yusun, Kang, Heyseung, & Park, Jieun. (2012). Is it vocabulary breadth or depth that better predicts Korean EFL learners' reading comprehension? *English Teaching*, *67*(4), 149–171. https://doi.org/10.15858/engtea.67.4.201212.149
Kelly, P. (1990). Guessing: No substitute for systematic learning of lexis. *System*, *18*(2), 199–207. https://doi.org/10.1016/0346-251X(90)90054-9
Kim, Myoungjin. (2019). Vocabulary size tests of different modality and their relationships with L2 reading and listening comprehension by Korean EFL learners in middle school. *Language Research*, *55*(1), 203–227. https://doi.org/10.30961/lr.2019.55.1.203
Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163
Landi, N., & Ryherd, K. (2017). Understanding specific reading comprehension deficit: A review. *Language and Linguistics Compass*, *11*(2), Article e12234. https://doi.org/10.1111/lnc3.12234
Laufer, B. (1989a). A factor of difficulty in vocabulary learning: Deceptive transparency. *AILA Review*, *6*, 10–20.
Laufer, B. (1989b). What percentage of text-lexis is essential for comprehension? In C. Laurén & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Multilingual Matters. https://doi.org/10.4324/9781003763703-16
Laufer, B. (1990). Why are some words more difficult than others? Some intralexical factors that affect the learning of words. *IRAL - International Review of Applied Linguistics in Language Teaching*, *28*(4), 293–308. https://doi.org/10.1515/iral.1990.28.4.293
Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). Palgrave Macmillan. https://doi.org/10.1007/978-1-349-12396-4_12
Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition: A rationale for pedagogy* (pp. 20–34). Cambridge University Press. https://doi.org/10.1017/CBO9781139524643.004
Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, *54*(4), 1076–1085. https://doi.org/10.1002/tesq.3004
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
Laufer-Dvorkin, B. (1985). *Vocabulary acquisition in a second language: The hypothesis of "synforms" (similar lexical forms)*

[Unpublished doctoral dissertation]. University of Edinburgh.
Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics*, *36(*5), 611–634. https://doi.org/10.1093/applin/amu007
Logan, J. K., & Kieffer, M. J. (2017). Evaluating the role of polysemous word knowledge in reading comprehension among bilingual adolescents. *Reading and Writing: An Interdisciplinary Journal*, *30*(8), 1687–1704. https://doi.org/10.1007/s11145-017-9745-1
McNeil, L. (2011). Investigating the contributions of background knowledge and reading comprehension strategies to L2 reading comprehension: An exploratory study. *Reading and Writing*, *24*(8), 883–902. https://doi.org/10.1007/s11145-010-9230-6
Ministry of Education. (2022). *Yeongeogwa gyoyuk gwajeong* [English curriculum] (Gyoyukbu gosi je 2022-33-ho [Ministry of Education Notice No. 2022-33, separate volume 14]).
Moon, R. (1997). Vocabulary connections: Multi-word items in English. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 40–63). Cambridge University Press.
Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178. https://doi.org/10.1037/h0027366
Nagy, W. E. (1995, November). *On the role of context in first- and second-language vocabulary learning* (Technical Report No. 627). Center for the Study of Reading, University of Illinois at Urbana-Champaign. https://hdl.handle.net/2142/31277
Nagy, W., Anderson, R. C., Schommer, M., Scott, J., & Stallman, A. C. (1989, January). *Morphological families in the internal lexicon* (Technical Report No. 450). Center for the Study of Reading, University of Illinois at Urbana-Champaign.
Nassaji, H. (2003a). Higher-level and lower-level text processing skills in advanced ESL reading comprehension. *The Modern Language Journal*, *87*(2), 261–276. https://doi.org/10.1111/1540-4781.00189
Nassaji, H. (2003b). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, *37*(4), 645–670. https://doi.org/10.2307/3588216
Nassaji, H. (2006). The relationship between depth of vocabulary knowledge and L2 learners' lexical inferencing strategy use and success. *The Modern Language Journal*, *90*(3), 387–401. https://doi.org/10.1111/j.1540-4781.2006.00431.x
Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759
Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59
Nation, P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9–13. https://doi.org/10.26686/wgtn.12552197.v1
Nurweni, A., & Read, J. (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes*, *18*(2), 161–175. https://doi.org/10.1016/S0889-4906(98)00005-2
Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, *98*(3), 554–566. https://doi.org/10.1037/0022-0663.98.3.554
Pellicer-Sánchez, A., Webb, S., & Wang, A. (2024). How does lexical coverage affect the processing of L2 texts? *Applied Linguistics*, *45*(6), 953–972. https://doi.org/10.1093/applin/amae062
Perfetti, C. A. (1986). Cognitive and linguistic components of reading ability. In B. R. Foorman & A. W. Siegel (Eds.), *Acquisition of reading skills: Cultural constraints and cognitive universals* (pp. 11–40). Erlbaum. https://doi.org/10.4324/9781315144412-2
Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730
Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). John Benjamins Publishing Company. https://doi.org/10.1075/swll.11.14per
Perfetti, C., Marron, M., & Foltz, P. (1996). Sources of comprehension failure: Theoretical perspectives and case studies. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 137–165). Erlbaum. https://doi.org/10.4324/9780203053324-7
Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687
Pigada, M., & Schmitt, N. (2006). Vocabulary acquisition from extensive reading: A case study. *Reading in a Foreign Language*, *18*(1), 1–28. https://doi.org/10.64152/10125/66611
Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Reading and Writing*, *25*(7), 1635–1664. https://doi.org/10.1007/s11145-011-9336-5
Qian, D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *The Canadian Modern Language Review*, *56*(2), 282–307. https://doi.org/10.3138/cmlr.56.2.282
Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*(3), 513–536. https://doi.org/10.1111/1467-9922.00193
Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, *11*(6), 1090–1098. https://doi.org/10.3758/BF03196742
Read, J. (1993). *The development of a new measure of L2 vocabulary knowledge*. Language Testing, *10*(3), 355–371. https://doi.org/10.1177/026553229301000308
Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment*

(pp. 41–60). Lawrence Erlbaum. https://doi.org/10.4324/9780203053768
Read, J. (2000). *Assessing vocabulary.* Cambridge University Press. https://doi.org/10.1017/CBO9780511732942
Read, J. (2004a). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 209–227). John Benjamins. https://doi.org/10.1075/lllt.10.15rea
Read, J. (2004b). Research in teaching vocabulary. *Annual Review of Applied Linguistics*, *24*, 146–161. https://doi.org/10.1017/S0267190504000078
Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, *12*(3), 329–363. https://doi.org/10.1177/1362168808089921
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
Seoul Metropolitan Office of Education. (2025). *2025nyeon 9wol go1·go2 jeonguk yeonhap hakryeok pyeongga tonggye jaryo·seongjeok bunseok jaryo* [Statistical and score analysis data for the September 2025 nationwide joint academic achievement assessment for first- and second-year high school students] [ZIP archive]. Retrieved October 23, 2025, from https://www.sen.go.kr/user/bbs/BD_selectBbsList.do?q_bbsSn=1036
Shen, M.-Y., & Wu, W.-S. (2009). Technical university EFL learners' reading proficiency and their lexical inference performance. *Electronic Journal of Foreign Language Teaching*, *6*(2), 189–200.
Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension* (MR-1465-OERI). RAND Corporation. http://www.jstor.org/stable/10.7249/mr1465oeri
Song, T., & Reynolds, B. L. (2022). The effect of lexical coverage on L2 learners' reading comprehension of narrative and expository genres. *Journal of English for Academic Purposes*, *59*, Article 101154. https://doi.org/10.1016/j.jeap.2022.101154
Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading, and writing. *Language Learning Journal*, *36*(2), 139–152. https://doi.org/10.1080/09571730802389975
Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading*, *10*(4), 381–398. https://doi.org/10.1207/s1532799xssr1004_3
Tunmer, W. E., & Chapman, J. W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities*, *45*(5), 453–466. https://doi.org/10.1177/0022219411432685
Tyson, R. (1993). English loanwords in Korean: Patterns of borrowing and semantic change. *EI Two Talk*, *1*(1), 29–36.
University of Oregon. (2023). *Dynamic Indicators of Basic Early Literacy Skills (DIBELS®) (8th ed.): Maze benchmark, Grade 6—Administration directions and scoring keys*. https://dibels.uoregon.edu
Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics*, *22*(2), 217–234. https://doi.org/10.1017/S0142716401002041
Webb, S. (2012). Depth of vocabulary knowledge. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1656–1663). Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal1325
Webb, S. (2021). Research investigating lexical coverage and lexical profiling: What we know, what we don't know, and what needs to be examined. *Reading in a Foreign Language*, *33*(2), 278–293. https://doi.org/10.64152/10125/67407
Widdowson, H. G. (1979). *Explorations in applied linguistics*. Oxford University Press.
Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: A depth of individual word knowledge model. *Studies in Second Language Acquisition*, *23*(1), 41–69. https://doi.org/10.1017/S0272263101001024
Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772
Year, Jung-eun. (2022). An exploratory study on vocabulary size of Korean high school EFL learners. *Modern English Education*, *23*(2), 22–32. https://doi.org/10.18095/meeso.2022.23.2.22
Zhang, D., & Koda, K. (2012). Contribution of morphological awareness and lexical inferencing ability to L2 vocabulary knowledge and reading comprehension among advanced EFL learners: Testing direct and indirect effects. *Reading and Writing*, *25*(5), 1195–1216. https://doi.org/10.1007/s11145-011-9313-z
Zhang, D., & Koda, K. (2017). Assessing L2 vocabulary depth with word associates format tests: Issues, findings, and suggestions. *Asian-Pacific Journal of Second and Foreign Language Education*, *2*, Article 1. https://doi.org/10.1186/s40862-017-0024-0
Zhang, D., & Yang, X. (2016). Chinese L2 learners' depth of vocabulary knowledge and its role in reading comprehension. *Foreign Language Annals*, *49*(4), 699–715. https://doi.org/10.1111/flan.12225

Appendix A

Summary of Measures, Scoring, and Reliability

Construct	Instrument	Format	Time (min)	Scoring	Reliability
Vocabulary Breadth	Vocabulary Size Test (VST; Nation & Beglar, 2007; online version via Lextutor)	MCQs; 30 items (1,000–3,000 word-family levels; 10 items per level)	15	1 point per item; 0–30	—
Vocabulary Depth	Word Associates Test (WAT; Read, 1993)	Select 4 of 8; 40 items	30	4 points per item; 0–160	—
Background Knowledge	Topic familiarity questionnaire (adapted from McNeil, 2011)	3-option format (True/False/Don't know); 6 items per topic, with 3 scored items used per topic	5	2 = correct; 1 = Don't know; 0 = incorrect; 0–12	α = .73
Reading Comprehension	Two adapted DIBELS 8th Ed. Maze passages, Grade 6 (University of Oregon, 2023)	10 MCQs per passage (7 factual, 2 inferential, 1 synthesis)	20 per passage	1 point per item; 0–10 per passage; CT RC and DT RC scored separately	KR-20: Form A = .72; Form B = .82
DT Sentence Interpretation	Sentence translation task (adapted from Laufer, 1989a, 2020)	5 sentence-based items from the DT passage read in the DT session	15	0–2 points per item; 0–10	Cohen's κ = .85 (20% subsample)

Appendix B

Representative Items from the Background Knowledge Questionnaire

Instruction: 3점 척도 ☐1 맞다. ☐2 아니다. ☐3 잘 모르겠다. 중 하나를 선택하세요.
1. 달의 표면은 걷기가 힘들 정도로 울퉁불퉁하다. ☐1 ☐2 ☐3
2. 심한 날씨 변화로 교통 수단이 멈추기도 한다. ☐1 ☐2 ☐3
3. 다른 나라의 특정한 날씨에 관한 글을 영어로 읽으면 이해할 수 있다. ☐1 ☐2 ☐3

Appendix C

Representative Items from the Vocabulary Size Test (VST)

Instruction: 총 30개의 문항이 있습니다. 각 예문에 쓰인 핵심 단어와 가장 가까운 뜻을 찾아서 체크하세요.

1. SEE: They saw it.
a. cut
b. waited for
c. looked at
d. started

2. TIME: They have a lot of time.
a. money
b. food
c. hours
d. friends

Appendix D

Representative Items from the Word Associates Test (WAT)

Instruction: **총 40문항으로 하나의 단어가 제시**되고 아래에는 8개의 단어가 주어집니다. 제시된 단어의 의미나 쓰임이 가장 밀접하게 관련된 **단어 4개를 선택해 주세요**. 관련된 단어들은 동의어, 함께 자주 쓰이는 단어(연어), 또는 같은 주제에 속하는 단어일 수 있습니다. **반드시 정확히 4개를 선택해야 합니다**. 추측에 대한 감점은 없지만, 최대한 정확하게 선택하려고 노력해주세요. 예시) 아래 1번에 제시된 단어 beautiful과 관련 있다고 생각되는 단어 4개를 찾아 체크하면 됩니다.

1. Beautiful

☐enjoyable ☐expensive ☐free ☐loud	☐education ☐face ☐music ☐weather

Appendix E

Representative Questions from the Reading Comprehension Test

CT1/DT1 Text Pair: Woodcut Printmaking

1. What is the oldest known method for making multiple copies of the same image? [Factual]
   A) Lithography
   B) Etching
   C) Woodcut
   D) Screen printing
   E) Photography
8. Why might the woodcut technique have spread to Europe around the same time as papermaking? [Inferential]
   A) Because books could only be printed after paper was available
   B) Because Europeans did not know how to carve wood before
   C) Because silk was too expensive to use for images in Europe
   D) Because printing presses were already common in Europe
   E) Because photographs had just been invented
10. Based on the passage, what combination of factors made woodcut printmaking an effective method for spreading knowledge in Europe? [Synthesis]
   A) The arrival of paper, the transfer of the technique, and the use of illustrations in books
   B) The invention of photography, the use of expensive materials, and the rise of universities
   C) The discovery of ink, the use of marble, and the growth of hardware stores
   D) The development of glass printing, the rise of posters, and the invention of silk fabric
   E) The growth of sculpture, the use of metal, and the popularity of clay tablets

CT2/DT2 Text Pair: Blizzards

3. Which regions in the United States are most commonly affected by blizzards? [Factual]
   A) The Southwest and West Coast
   B) The Great Plains, the Great Lakes, and the Northeast
   C) The Southeast and Florida
   D) The Pacific Northwest and Alaska
   E) The Rocky Mountains and the Desert Southwest
9. Why do drivers try to stay off the roads during a blizzard? [Inferential]
   A) Because roads become slippery with rain
   B) Because the wind blows cars off the road
   C) Because windblown snow makes it hard to see
   D) Because roads are closed for construction
   E) Because blizzards only happen at night
10. What is the main idea of the passage about blizzards? [Synthesis]
   A) Blizzards are light snow showers that cause little harm.
   B) Blizzards are historic storms that can seriously disrupt daily life.
   C) Blizzards are fun events for children to play in.
   D) Blizzards only happen in Florida and Alabama.
   E) Blizzards are short storms that last only a few hours.

Appendix F

Representative Items from the DT Sentence Interpretation Task

Scoring Rubric (0–2)
2 = Fully correct (Accurate contextual meaning conveyed; appropriate sense and function)
1 = Partially correct (Gist captured but sense incomplete or awkward; minor semantic and functional errors)
0 = Incorrect (Dominant but wrong sense; unrelated meaning; blank or irrelevant response)

Instruction: 다음 문장을 한국어로 해석해 쓰세요.

DT1 Sample Sentences
1. Some artists became famous for their single-**leaf** prints.
2. Then you spread it on a piece of glass or marble, and roll it out in a thin, **even** layer with a hard rubber roller.

DT2 Sample Sentences
1. Blizzard is **a term** used to describe a kind of massive storm that completely covers everything with snow.
2. Snow may cover **barns** and other places where food is grown or stored.