Article

Not peer-reviewed version

# Balanced Efficiency in EFL Digital Listening-Reading: A Biopsychological Model of Cognitive Load, Engagement, and Lexical Retention

Heeseong Ahn and Chungwan Lim *

Posted Date: 6 April 2026

doi: 10.20944/preprints202604.0365.v1

Keywords: cognitive load; multimodal input; second language acquisition

*Article*

# Balanced Efficiency in EFL Digital Listening-Reading: A Biopsychological Model of Cognitive Load, Engagement, and Lexical Retention

**Heeseong Ahn [1] and Chungwan Lim [2,*]**

[1] Center for In-Service Education, Korean National University of Education, Cheongju, Chungbuk 28173, Republic of Korea

[2] Department of Earth Science Education, Kongju National University, Kongju 32588, Republic of Korea

* Correspondence: tephra@kongju.ac.kr

**Abstract**

In second language learning, research on multimodal input often assumes that when learners receive both audio and written text, comprehension becomes easier because cognitive load is reduced. This assumption, however, may not fully explain how multimodal input works in digital reading. The present study reexamines this assumption by investigating whether listening-reading input regulates learner engagement rather than simply lowering cognitive demand during digital EFL reading. Forty Korean university EFL learners were randomly assigned to either a text-only reading condition or a listening-reading condition in which audio accompanied the written text. Vocabulary knowledge was measured using a 20-item test administered before and after the intervention, and delayed retention was examined descriptively. Heart rate data were continuously recorded in order to examine physiological responses during task performance and the recovery period. The results showed that the listening–reading group made clear improvement in vocabulary scores ($M = 67.00 \rightarrow 80.25$; $t(19) = -11.395$, $p < .001$, $\eta^2 = .872$, $d = 2.47$). In contrast, the text-only group did not show statistically significant improvement ($p = .096$). Contrary to the expectation that multimodal input would reduce physiological load, participants in the listening-reading condition exhibited higher task heart rate ($d = .59$) and greater elevation relative to baseline ($d = .63$). These results indicate that listening-reading may facilitate lexical acquisition through heightened engagement and attentional activation rather than simple cognitive load reduction. The findings provide additional insight for research on multimodal input in SLA and suggest that an appropriate level of cognitive engagement may play an important role in digital reading environments.

**Keywords:** cognitive load; multimodal input; second language acquisition

## 1. Introduction

Vocabulary knowledge plays an essential role in second language (L2) reading comprehension and overall language proficiency, particularly in English as a Foreign Language (EFL) contexts where reading functions as a primary source of linguistic input [1,2]. Lexical knowledge not only predicts reading comprehension but also supports listening comprehension, academic discourse participation, and long-term language attainment [3,4]. From a usage-based perspective, repeated exposure and attention to language input allows connections between word forms and meanings to become stronger over time [5,6]. Therefore, instructional methods that influence how learners process language input may have significant effects on vocabulary development, even when the amount of input remains similar.

At the same time, digital learning environments have significantly changed how learners encounter language input. Contemporary reading platforms now combine written texts with additional features such as synchronized audio narration, adjustable pacing, glosses, and adaptive

 doi:10.20944/preprints202604.0365.v1



learning tools. Such multimodal systems raise a central theoretical and design question: do they indeed enhance learning primarily by reducing cognitive load, by amplifying learner engagement, or by simultaneously calibrating both processes [7–9]? Although multimedia learning theory provides useful principles for combining visual and auditory information, second language acquisition (SLA) research has not yet fully explained how these principles influence vocabulary learning in realistic classroom settings.

A common assumption among educators is that listening while reading makes reading easier for language learners. Audio support may assist learners in recognizing pronunciation patterns, mapping written forms to sounds, and identifying prosodic cues in the text and thereby lower processing strain. According to Cognitive Load Theory, lowering unnecessary processing demands, often referred to as extraneous load, allows learners to allocate more cognitive resources to meaningful learning processes such as comprehension and schema formation [10]. Within L2 contexts, this may facilitate lexical encoding by reallocating cognitive resources toward semantic integration rather than surface-level decoding.

However, recent research in learning sciences suggests that this explanation may be too simple. Many scholars now distinguish between cognitive load and learner engagement as related but separate aspects of the learning process [11–13]. Engagement-based frameworks often conceptualize learning as involving cognitive effort, behavioral participation, and emotional involvement [14,15]. Under this perspective, effective instruction may reduce certain processing difficulties while at the same time increasing sustained attention, motivation, and task involvement. In such cases, learning activities might not appear less demanding from a physiological perspective; instead, they may produce higher levels of attentional activation while still supporting learning.

This distinction carries methodological and theoretical significance for SLA research incorporating psychophysiological measures. Physiological indices such as heart rate (HR) and heart rate variability (HRV) are sensitive to integrated autonomic activation reflecting both cognitive effort and affective arousal [16,17]. Elevated HR may signal increased processing demand, heightened attentional mobilization, emotional engagement, or a combination thereof. Without theoretical grounding and behavioral triangulation, interpreting HR elevation as unequivocal evidence of “higher load” risks conflating adaptive engagement with detrimental overload. In digital language learning contexts, such misinterpretation could lead to overly simplified instructional designs aimed at minimizing activation rather than optimizing learning.

Recent SLA discussions call for greater integration of process measures with outcome data to strengthen mechanistic explanations [18–20]. Technologies such as eye-tracking and pupillometry have been used to examine attention during language processing [18], comparatively few studies have examined autonomic indicators such as HR or HRV in vocabulary learning contexts. Even fewer have explicitly modeled the interplay between multimodal input, cognitive load, engagement, and physiological regulation. As a result, the field lacks a coherent biopsychological account of how digital listening-reading environments reshape lexical processing states.

The present study attempts to address this issue by considering listening-while-reading as a mechanism that may balance cognitive efficiency and learner engagement. When written text is combined with synchronized audio, decoding difficulties may decrease, which can improve processing efficiency. At the same time, the presence of audio may maintain learner attention and encourage sustained task involvement, potentially leading to higher levels of physiological activation. Rather than assuming that effective multimodal learning must reduce physiological arousal, this study investigates whether vocabulary learning is associated with reduced strain, increased engagement-related activation, or a balanced combination of both.

Using experimental data comparing text-only and listening-while-reading conditions among Korean university EFL learners, we investigate two core questions: (a) whether multimodal listening-reading produces superior vocabulary acquisition and retention, and (b) whether heart rate patterns during task performance support a cognitive-load reduction explanation, an engagement-based explanation, or a balanced interpretation involving both processes. By integrating behavioral outcomes with physiological measures, this study contributes to a theoretically grounded model of cognitive-affective regulation in digital L2 reading environments.

Accordingly, the present study has three main purposes:

1. To compare vocabulary learning outcomes under text-only and text-with-audio listening-while-reading conditions.
2. To examine heart rate dynamics during task performance and recovery as indicators of cognitive-affective regulation.
3. To evaluate whether multimodal digital reading operates under a reduced-strain profile, an engagement-amplification profile, or a calibrated "balanced efficiency" profile.

By examining listening-while-reading within a biopsychological framework that integrates cognitive load theory, learner engagement, and autonomic regulation, this study seeks to advance a more mechanistic understanding of digital language learning in EFL contexts.

## 2. Theoretical Background

### *2.1. Multimodal Input, Dual Coding, and Lexical Encoding*

Dual Coding Theory posits that information encoded through verbal and nonverbal channels is represented in partially independent yet interconnected systems, increasing the number of potential retrieval pathways and enhancing memory stability [21]. When learners process coordinated auditory and visual input, encoding may benefit from representational redundancy, facilitating later access through multiple associative routes. Multimedia learning theory builds on this idea and suggests that learning improves when information is distributed across modalities in ways that reduce redundancy, maintain coherence, and align temporal presentation [7,8].

Within L2 vocabulary learning, lexical acquisition involves establishing stable connections among orthographic form, phonological representation, semantic meaning, and contextual usage [1,2]. From a usage-based perspective, lexical representations strengthen through repeated activation and contextualized processing [5,6]. Through repeated exposure and attention, the relationships between form and meaning become stronger. Synchronized audio during reading may enhance phonological encoding, support prosodic parsing, and reduce uncertainty in grapheme-phoneme mapping, thereby promoting more precise lexical binding.

Listening-while-reading may therefore improve lexical consolidation not only by increasing input quantity but by altering the depth and quality of encoding. Studies related to the production effect and enriched encoding also show that when learners process information through multiple modalities, the representations may become more distinctive in memory [3,22]. These findings collectively suggest that synchronized text-audio presentation may strengthen phonological-orthographic integration and support long-term lexical accessibility.

However, the advantages of multimodal input do not occur automatically. Cognitive Load Theory indicates that poorly designed multimedia presentations can induce split attention or redundancy effects, thereby impairing learning [7,12]. Thus, listening-while-reading effectiveness depends on coordinated alignment rather than mere modality addition. In digital environments, adaptive pacing and synchronized narration may mitigate such risks, but the cognitive-affective consequences of these features remain underexplored in SLA contexts.

### *2.2. Cognitive Load and "Balanced Efficiency" in Digital Reading*

Cognitive Load Theory (CLT) explains how limitations of working memory influence learning processes. The theory distinguishes among intrinsic load (task complexity inherent to the material), extraneous load (demands imposed by suboptimal presentation), and germane load (cognitive resources devoted to schema construction) [9,10]. In L2 reading, intrinsic load may arise from unfamiliar vocabulary, syntactic complexity, or limited automaticity. Extraneous load may stem from

doi:10.20944/preprints202604.0365.v1



decoding inefficiencies, pacing difficulties, or fragmented presentation formats. Germane load reflects the learner's investment in integrating lexical items into existing semantic networks.

Listening-while-reading has the potential to reduce extraneous load by easing decoding demands and supporting phonological processing, thereby freeing cognitive resources for semantic integration [1,7]. In listening-while-reading situations, the presence of synchronized audio may help reduce certain types of extraneous load. When learners hear the pronunciation of the text while reading, they may experience fewer difficulties in decoding written forms and identifying phonological patterns. As a result, some cognitive resources may become available for deeper semantic processing and vocabulary integration. However, CLT also suggests that effective learning does not always occur when effort is minimized. In many cases, meaningful learning requires learners to invest cognitive effort in processing and integrating new information [12]. This effort corresponds to germane load, which supports the construction of knowledge structures.

Recent SLA studies increasingly recognize that learning is not solely determined by cognitive capacity but by dynamic regulation of attention and engagement [11,14]. Engagement frameworks often describe learning as involving several components, including cognitive effort, emotional involvement, and behavioral participation [15]. Within this perspective, effective digital reading environments may not minimize activation but calibrate it, sustaining attentional engagement while avoiding overload.

From this perspective, effective digital reading environments may not simply aim to reduce mental effort. Instead, they may create conditions where learners maintain sustained attention without experiencing excessive cognitive strain. In other words, successful learning environments may regulate the level of activation so that learners remain engaged while still processing the material effectively.

In the present study, this idea is described as balanced efficiency. Balanced efficiency refers to a learning condition in which unnecessary processing demands are reduced while meaningful cognitive effort is maintained. In this situation, physiological activation may reflect focused attention and engagement rather than negative overload: an instructional state in which extraneous load is reduced, germane processing is enhanced, and physiological activation reflects engaged attention rather than maladaptive strain. In digitally mediated contexts, features such as synchronized narration, adaptive pacing, and dynamic highlighting may facilitate this balance by coordinating perceptual input with attentional rhythms [7,12]. However, empirical validation of this theoretical construct requires integrating behavioral and physiological evidence.

### *2.3. Physiological Indices: Interpreting Heart Rate Dynamics*

Physiological measurements can provide useful information about cognitive and emotional processes that occur during learning activities. Heart rate (HR) and heart-rate variability (HRV) reflect the integrated output of sympathetic and parasympathetic nervous system activity [16,17]. Changes in heart rate during a task may therefore indicate several processes, such as increased cognitive effort, emotional arousal, motivational involvement, or stress-related activation. HRV measures, including indices such as root mean square of successive differences (RMSSD) and high-frequency power, are often interpreted as indicators of regulatory control and flexibility in physiological responses.

Importantly, elevated HR does not inherently indicate harmful overload. Educational psychology research demonstrates that moderate physiological activation often accompanies sustained attention and motivated task involvement [12,13]. Within L2 contexts, heightened activation during meaningful processing may reflect increased lexical integration effort rather than cognitive breakdown. Conversely, low physiological activation may signal disengagement rather than optimal efficiency.

In SLA research, physiological measures have been less frequently employed compared to behavioral indices or self-reported cognitive load [18,19]. When physiological indicators are included, they are often interpreted unidimensionally as indicators of "load," overlooking their

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2026 doi:10.20944/preprints202604.0365.v1



sensitivity to affective and regulatory processes. For this reason, contemporary perspectives emphasize the need to triangulate physiological data with learning outcomes to avoid misattributing engagement-related activation to overload [20].

Following this perspective, the present study interprets heart-rate dynamics within a theoretically integrated framework that considers cognitive load, engagement, and retention outcomes simultaneously. Instead of assuming that physiological activation always represents negative overload, we examine whether changes in heart rate occur together with improved vocabulary learning. If higher activation is accompanied by better learning performance, it may indicate productive engagement within a balanced efficiency condition. In contrast, if increased activation occurs without learning improvement, it may suggest excessive processing difficulty.

By integrating multimodal input theory, cognitive load distinctions, engagement research, and psychophysiological measurement, this framework advances a biopsychological account of digital listening-reading in SLA. The following sections empirically test whether vocabulary gains under text-with-audio conditions align with reduced-strain, engagement-amplification, or balanced efficiency interpretations.

## 3. Methods

### *3.1. Participants and Design*

Participants were 40 Korean university students enrolled in required English courses at a large metropolitan university. All participants were native speakers of Korean and had received a minimum of six years of prior formal English instruction through secondary education. Their proficiency level corresponded approximately to lower-intermediate to intermediate based on institutional placement criteria. None reported hearing impairments or cardiovascular conditions that would affect physiological measurement.

Participants were randomly assigned to one of two experimental conditions:

- Text-only (reading-only) condition (n = 20)
- Text-with-audio (listening-while-reading) condition (n = 20)

The study employed a between-subjects pretest-posttest experimental design with physiological monitoring during the learning phase (see Figure 1). Random assignment was conducted to minimize systematic group differences. Pretest vocabulary equivalence was statistically verified prior to outcome analyses.

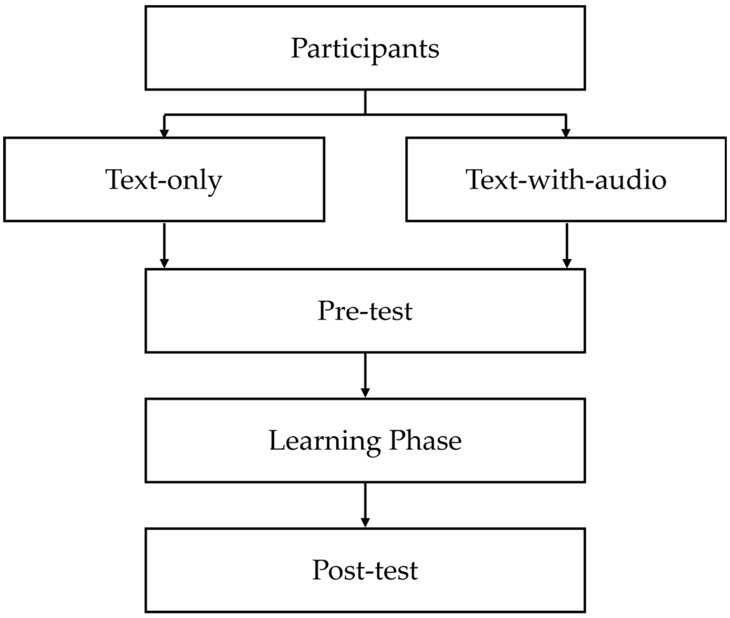


**Figure 1.** Methodology flowchart of the participant-related phase.

The learning task consisted of a controlled digital reading activity containing target lexical items embedded in short expository passages. The text-with-audio condition received synchronized auditory narration aligned with the written text. The text-only group viewed identical textual content without audio support. The duration of the learning phase was standardized across conditions to ensure equal exposure time.

Physiological data were recorded continuously during:

1. A baseline period (resting state prior to task)
2. A 120-second task window
3. A post-task recovery phase

This design enabled simultaneous examination of behavioral learning outcomes and cognitive-affective activation patterns.

### *3.2. Measures*

#### 3.2.1. Vocabulary Knowledge

Vocabulary learning was assessed using a 20-item researcher-developed test, with each item scored on a five-point scale (maximum total score = 100). Items targeted lexical form-meaning mapping for words embedded in the reading passages. The test was administered (1) immediately prior to the intervention (pretest); and (2) immediately after the intervention (posttest).

Items were presented in multiple-choice format with four alternatives per item. Scoring procedures were standardized across groups. Internal consistency reliability (Cronbach's $\alpha$) was examined to ensure acceptable measurement stability. Pretest scores were used to confirm group equivalence, while posttest scores were used to evaluate learning gains. Descriptive statistics (means and standard deviations) were calculated for both time points.

#### 3.2.2. Physiological Measures

Heart rate (HR) was recorded using a wearable monitoring device validated for research-grade task-based measurement. Data were collected during three phases:

1. **Baseline phase**: resting HR recorded prior to task onset
2. **Task phase:** 120-second reading activity
3. **Recovery phase**: post-task resting period

Heart-rate values were sampled at 10-second intervals during the task phase, resulting in 12 time points per participant. From these time-series data, four derived indices were calculated:

1. **Mean Task Heart Rate** (MTHR):
   The average HR across the 120-second task window.
2. **Mean Task Elevation Over Baseline** (ΔHR):
   MTHR minus baseline HR, indexing task-related activation.
3. **Peak Task Heart Rate**:
   The highest HR value recorded during the task window.
4. **Recovery Time**:
   The number of seconds required for HR to return toward baseline following task completion.

Baseline normalization procedures were applied to account for individual variability in resting heart rate. Physiological indices were interpreted within a cognitive-affective regulation framework rather than as direct proxies for cognitive load.

 doi:10.20944/preprints202604.0365.v1



### *3.3. Data Analysis Measures*

#### 3.3.1. Preliminary Analyses

Group equivalence at pretest was evaluated using an independent-samples t-test. Assumptions of normality and homogeneity of variance were examined prior to inferential testing. Descriptive statistics were computed for all behavioral and physiological measures. Outliers were screened using standard deviation criteria and visual inspection of boxplots.

#### 3.3.2. Behavioral Analyses

Within-group vocabulary gains (pretest vs. posttest) were examined using paired-samples t-tests for each condition separately. Between-group comparisons at posttest were conducted using independent-samples t-tests.

Effect sizes were reported following best practices in quantitative SLA research:

- Eta squared ($\eta^2$) for within-group comparisons
- Cohen's d for between-group comparisons

Effect sizes were interpreted using established conventions [19,23], while also acknowledging field-specific benchmarks in applied linguistics.

Group equivalence at pretest was evaluated using an independent-samples t-test. Assumptions of normality and homogeneity of variance were examined prior to inferential testing. Descriptive statistics were computed for all behavioral and physiological measures. Outliers were screened using standard deviation criteria and visual inspection of boxplots.

#### 3.3.3. Physiological Analyses

Physiological indices (MTHR, ΔHR, peak task HR, recovery time) were compared between conditions using independent-samples t-tests.

Cohen's d was reported to quantify the magnitude of between-group differences. Because physiological activation may reflect both cognitive effort and engagement-related arousal, results were interpreted in conjunction with behavioral learning outcomes.

Specifically, elevated physiological activation paired with superior vocabulary gains was interpreted as evidence of adaptive engagement rather than maladaptive overload.

All analyses were conducted using standard statistical software. Statistical significance was set at $\alpha = .05$.

## 4. Results

### *4.1. Vocabulary Outcomes*

#### 4.1.1. Baseline Equivalence

An independent-samples t-test was conducted to examine group equivalence at pretest (see Table 1). The text-with-audio group (M = 67.00, SD = 4.10) and the text-only group (M = 66.75, SD = 3.73) did not differ significantly, $t(38) = −0.202$, $p = .841$, $\eta^2 = .001$. The effect size was negligible, confirming that any subsequent differences in learning outcomes cannot be attributed to initial vocabulary disparities. Assumptions of homogeneity of variance were met, and the distribution of scores did not deviate substantially from normality.

 doi:10.20944/preprints202604.0365.v1



**Table 1.** Vocabulary Scores by Condition (Pretest and Posttest).

| Condition | N | Pre M | Pre SD | Post M | Post SD |
|---|---|---|---|---|---|
| Text-only | 20 | 66.75 | 3.73 | 68.00 | 4.70 |
| Text-with-Audio | 20 | 67.00 | 4.10 | 80.25 | 6.38 |

*Note: Text-with-Audio gain: $\eta^2 = .872$, d = 2.47. Text-only gain: $\eta^2 = .139$, d = 0.29.

4.1.2. Within-Group Learning Gains

Paired-samples t-tests were conducted separately for each condition. The text-with-audio group demonstrated a substantial improvement from pretest (M = 67.00, SD = 4.10) to posttest (M = 80.25, SD = 6.38), t(19) = 11.395, p < .001. The effect size was extremely large ($\eta^2 = .872$; Cohen's d = 2.47) (see Table 1), indicating that approximately 87% of the variance in gain scores within this group was attributable to the intervention.

In contrast, the text-only group showed only a modest numerical increase from pretest (M = 66.75, SD = 3.73) to posttest (M = 68.00, SD = 4.70), which did not reach statistical significance (t(19) = −1.751, p = .096) (see Table 1). The associated effect size was small ($\eta^2 = .139$; d = 0.29). These results indicate a clear divergence in learning trajectories across conditions: while listening-reading produced robust vocabulary gains, text-only reading resulted in minimal improvement over the same exposure duration.

4.1.3. Between-Group Posttest Comparison

Although groups were equivalent at pretest, posttest scores diverged substantially. The magnitude of the difference in gain scores suggests that synchronized audio contributed meaningfully to lexical development beyond exposure to written input alone. The extremely large within-group effect size for the text-with-audio condition (d = 2.47) exceeds typical benchmarks reported in L2 instructional research [16], underscoring the practical as well as statistical significance of the intervention.

*4.2. Physiological Outcomes*

4.2.1. Task-Period Activation

Contrary to a strict reduced-strain interpretation, physiological indices during the 120-second task window tended to be higher in the text-with-audio condition. Mean task heart rate (MTHR) was higher for the text-with-audio group (M = 76.34 bpm, SD = 4.63) than for the text-only group (M = 72.91 bpm, SD = 6.75) (t(38) = 1.872, p = .069, Cohen's d = 0.59) (see Table 2). Although this difference did not reach conventional significance at $\alpha = .05$, the effect size falls within the moderate range.

Similarly, mean task elevation above baseline (ΔHR) was greater for text-with-audio (M = 6.84 bpm, SD = 4.72) than for text-only (M = 3.11 bpm, SD = 6.90) (t(38) = 1.994, p = .053, d = 0.63) (see Table 2). The direction and magnitude of this effect again indicate moderately higher activation in the listening-reading condition. Importantly, these differences were consistent in direction across activation indices, suggesting a systematic pattern rather than random fluctuation.

**Table 2.** Heart Rate Indices by Condition (Task Window and Recovery).

| Index | Text-with-Audio M | Text-with-Audio SD | Text-only M | Text-only SD | t(38) | p | Cohen's d |
|---|---|---|---|---|---|---|---|
| MTHR (bpm) | 76.34 | 4.63 | 72.91 | 6.75 | 1.872 | .069 | 0.59 |
| ΔHR (bpm) | 6.84 | 4.72 | 3.11 | 6.90 | 1.994 | .053 | 0.63 |
| Recovery time (s) | 116.55 | 40.51 | 109.40 | 32.01 | 0.619 | .539 | 0.20 |
| Peak task HR (bpm) | 84.95 | 3.83 | 84.35 | 3.91 | 0.490 | .627 | 0.15 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Peak elevation (bpm) | 15.45 | 2.96 | 14.55 | 2.84 | 0.981 | .333 | 0.31 |

* Note: Positive d values indicate higher values for Text-with-Audio relative to Text-only.

4.2.2. Peak Activation and Recovery

Peak task heart rate did not differ substantially between groups: text-with-audio (M = 84.95 bpm, SD = 3.83) and text-only (M = 84.35 bpm, SD = 3.91) (t(38) = 0.490, p = .627, d = 0.15) (see Table 2). Similarly, peak elevation above baseline showed only a small difference (d = 0.31), and recovery time following task completion did not differ significantly (t(38) = 0.619, p = .539, d = 0.20).

The absence of pronounced differences in peak values or recovery duration suggests that the higher average task activation in the text-with-audio condition did not translate into prolonged physiological dysregulation.

5.2.3. Integrative Pattern

Taken together, the physiological results reveal:

- Moderately higher sustained activation during task performance in the text-with-audio condition
- No evidence of extreme peak differences
- No significant prolongation of recovery

This profile is inconsistent with a simple overload interpretation and instead suggests a pattern of sustained, moderate activation during listening-reading.

## 5. Discussion

### *5.1. Consistency with theoretical perspectives*

The main contribution of the present study is conceptual rather than merely empirical. The findings support a balanced efficiency interpretation of multimodal digital reading wherein improved vocabulary learning co-occurs with moderate physiological activation. This pattern suggests that listening-while-reading may not simply reduce cognitive load, but may instead adjust the cognitive and affective conditions under which learning takes place.

The vocabulary results are consistent with theoretical perspectives such as Dual Coding Theory [21] and multimedia learning principles [7,8], in which the text-with-audio condition yielded very large vocabulary gains. When auditory and visual information are presented together, learners may receive complementary cues that support phonological and orthographic processing. As a result, the connections between word forms and meanings may become stronger. From a usage-based perspective, such synchronized input may increase the salience and stability of lexical form-meaning mappings [5,6]. The magnitude of the behavioral effect suggests that the multimodal configuration substantially altered encoding conditions rather than simply increasing exposure.

At the same time, physiological evidence complicates a straightforward cognitive load reduction account. Mean task heart rate and mean elevation over baseline were higher in the text-with-audio condition, with moderate effect sizes. These patterns are not consistent with a pure reduced-strain model in which effective instruction should correspond to lower physiological activation. Instead, interpreted through a cognitive-affective framework, the findings align more closely with engagement-related arousal [12,13]. That is, learners may have mobilized greater attentional resources when presented with synchronized auditory and textual input.

It is important to note that physiological activation cannot be equated with inefficiency. Heart-rate dynamics reflect integrated autonomic responses shaped by cognitive effort, affective engagement, and regulatory control [17]. Moderate increases in heart rate during learning tasks may indicate sustained attention rather than excessive strain. The absence of prolonged recovery time differences further supports this interpretation: although activation was higher during the task, it did not produce evidence of extended physiological dysregulation.

Crucially, the behavioral results provide an external validity check on physiological interpretation. If the higher activation levels represented harmful cognitive overload, poorer learning outcomes would likely be observed in the listening-reading condition. Instead, the opposite pattern emerged: the group exhibiting higher activation achieved dramatically greater vocabulary gains. This dissociation undermines the assumption that lower physiological activation necessarily corresponds to more effective learning [9]. Rather, it suggests that optimal learning conditions may involve calibrated activation that sustains goal-directed engagement.

From the perspective of Cognitive Load Theory, listening-reading may reduce extraneous load by easing decoding demands and supporting phonological parsing. Simultaneously, it may increase germane load by encouraging deeper lexical integration and sustained attention [10]. In this formulation, improved outcomes emerge not from minimizing effort but from redistributing cognitive resources toward meaningful processing. The richer multimodal stimulus environment may heighten attentional focus, thereby increasing arousal indices while enhancing encoding quality.

This interpretation is also consistent with engagement-based frameworks in SLA, which conceptualize effective learning as arising from coordinated cognitive, emotional, and behavioral involvement [11,14]. Listening-reading may support such coordinated engagement by aligning auditory and visual channels, reducing ambiguity, and sustaining attention through temporal synchronization. Under these conditions, elevated heart-rate indices in this context may reflect motivated task involvement rather than stress-related strain.

The findings carry broader implications for how researchers conceptualize and operationalize "cognitive load" in technology-mediated SLA. Physiological indicators such as HR should not be treated as unidimensional indicators of processing difficulty. Instead, they must be interpreted within a multi-component framework that acknowledges interactions among cognitive demand, affective arousal, and regulatory mechanisms [12,17]. Without behavioral triangulation, physiological elevation risks being misclassified as overload, potentially leading to overly conservative instructional designs aimed at minimizing activation rather than optimizing learning.

For digital language learning systems, the design goal should therefore not be the elimination of physiological activation but its calibration. Optimal environments may support sustained attentional mobilization while preventing cognitive saturation. The present data suggest that synchronized listening-reading can achieve such calibration: vocabulary learning improved substantially even as arousal indices increased. In other words, increased activation did not signal breakdown but adaptive engagement.

More broadly, this study contributes to an emerging biopsychological approach to SLA that integrates behavioral outcomes with process-level indicators. By examining vocabulary gains alongside physiological dynamics, the findings demonstrate that multimodal digital reading reshapes the cognitive-affective ecology of learning. Future research incorporating HRV, eye-tracking, or longitudinal retention modeling could further clarify how engagement and regulation interact over time.

All in all, the results challenge the assumption that effective instruction must always appear cognitively or physiologically easier. Instead, the findings support a model of balanced efficiency in which multimodal listening-reading reduces extraneous barriers while sustaining productive engagement. Within this framework, moderate increases in physiological activation may not indicate overload. When they occur together with strong learning outcomes and stable recovery patterns, they may represent an adaptive state that supports durable vocabulary development.

 doi:10.20944/preprints202604.0365.v1



### 5.2. Implications for SLA Theory

The present findings have implications for theoretical models of second language acquisition that conceptualize learning as emerging from interactions among input properties, attentional allocation, affective engagement, and cognitive regulation. Robinson's [24] Cognition Hypothesis and Schumann's [25] neurobiological perspective both emphasize that learning is shaped not only by linguistic input but also by learners' attentional and affective states during processing. The current results align with such integrative models by demonstrating that vocabulary gains depend on how multimodal input reorganizes cognitive-affective activation rather than merely reducing processing difficulty.

In particular, the balanced efficiency account extends Cognitive Load Theory [9,10] by foregrounding the interaction between load and engagement. While CLT traditionally emphasizes minimizing extraneous load to protect limited working memory capacity, the present findings suggest that productive learning may occur under moderate physiological activation when that activation reflects sustained attentional engagement rather than inefficiency. This view resonates with Paas and Ayres' [12] broader formulation of load theory as a dynamic allocation of cognitive resources, as well as with contemporary engagement frameworks in SLA that conceptualize learning as a coordinated cognitive-affective process [11,14].

From the perspective of usage-based theory and complex dynamic systems approaches [6,26], listening-reading may reshape the temporal dynamics of lexical activation by increasing salience and strengthening form-meaning integration. Elevated physiological activation during multimodal input could reflect heightened attentional investment that accelerates the stabilization of lexical representations. Rather than conceptualizing learning as occurring under minimal activation, these findings support models in which optimal development occurs within calibrated activation zones-states in which cognitive effort and affective engagement are balanced.

Furthermore, the study contributes methodologically to SLA by illustrating the value of integrating behavioral outcomes with psychophysiological measures. As calls for process-oriented research intensify [18,19], the inclusion of heart-rate indices demonstrates how physiological data can refine interpretations of instructional effectiveness. At the same time, the results indicate that physiological indicators should not be interpreted in a simple way. Lower physiological activation does not necessarily indicate more effective learning conditions.

In summary, the findings support theoretical approaches that view second language acquisition as a multi-component process in which linguistic input, attentional processes, emotional engagement, and regulatory mechanisms interact during learning. Understanding these interactions may help researchers better explain how different instructional conditions influence language development.

### *5.3. Implications for Educational Practice*

For educational practice, particularly in digitally mediated language learning, the findings suggest that multimodal design should be conceptualized as a resource-allocation tool rather than merely a load-reduction device. In many cases, multimodal instructional design has been understood mainly as a way to reduce cognitive load. However, the results of this study suggest that multimodal features may function more as tools for allocating cognitive resources during learning. In digital reading platforms, design elements such as synchronized audio narration, adjustable pacing, and interactive glosses may help learners focus attention on lexical processing while maintaining engagement with the reading task.

For developers of digital language learning systems, including those incorporating adaptive or AI-guided features, should view synchronized audio, pacing controls, and dynamic glossing as mechanisms for reallocating cognitive resources toward lexical integration while sustaining attentional engagement [1,7].

Crucially, platform evaluation should move beyond accuracy metrics alone. Indicators of engagement, such as time-on-task, voluntary rereading behavior, interaction patterns, or even physiological proxies, may provide complementary insight into whether activation is productive or

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2026 doi:10.20944/preprints202604.0365.v1



distracting [11,13]. immediately be interpreted as a problem that requires simplification of the instructional design. Instead, developers should assess whether activation co-occurs with improved retention and comprehension.

Teachers implementing listening-reading activities in classroom settings may similarly benefit from recognizing that moderate task-related arousal can be pedagogically desirable. Rather than aiming to make tasks uniformly "easy," instructors might focus on creating multimodal environments that reduce unnecessary decoding strain while maintaining attentional challenge. In practice, this may involve synchronizing audio with text for unfamiliar vocabulary, encouraging controlled pacing, or strategically embedding glosses to support semantic integration without fragmenting attention.

Overall, these findings suggest that technology-enhanced language learning should not focus only on reducing cognitive demand. A more effective approach may involve maintaining an appropriate balance between cognitive effort and learner engagement during the learning process.

## 6. Limitations and Future Research

Several limitations of the present study should be acknowledged.

First, the sample size (n = 20 per condition) limits statistical precision, particularly for detecting smaller physiological differences. While large behavioral effect sizes mitigate concerns regarding power for vocabulary outcomes [23,27], future research should employ larger samples to test potential moderating variables such as proficiency level, working memory capacity, or language anxiety [5,24]. Such analyses could clarify whether balanced efficiency effects generalize across learner profiles.

Second, heart-rate measures were used as indicators of task-related physiological activation. While HR reflects overall autonomic nervous system activity, it cannot clearly separate cognitive effort from emotional arousal or stress responses. For this reason, future research should incorporate additional measurements to better interpret cognitive–affective states during learning. Possible measures include heart-rate variability (HRV), subjective workload instruments such as the NASA Task Load Index (NASA TLX), and behavioral indicators of engagement. Using multiple measurement methods may help researchers more accurately determine whether increased physiological activation reflects productive engagement or excessive cognitive strain.

Third, delayed retention was examined descriptively rather than modeled longitudinally. Durable lexical development is best evaluated through repeated measurement across extended intervals [1,3]. Future research should apply latent growth modeling or mixed-effects approaches to track retention trajectories over weeks or months. Such modeling would allow researchers to determine whether multimodal activation effects translate into sustained lexical consolidation.

Fourth, the experimental design of the present study focused on short-term processing during controlled reading tasks. Additional research is needed to examine whether listening-reading influences other aspects of language performance, such as reading fluency, comprehension monitoring, or transfer to new reading materials. Process-oriented methods, including eye-tracking or pupillometry, may also provide more detailed information about attentional allocation during multimodal input processing [18].

Finally, expanding this research across diverse linguistic backgrounds and instructional contexts would enhance generalizability. Cross-linguistic comparisons could determine whether phonological transparency or orthographic distance moderates the benefits of synchronized audio. The present findings have implications for theoretical models of second language acquisition that conceptualize learning as emerging from interactions among input properties, attentional allocation, affective engagement, and cognitive regulation [24] (see Figure 2).

 doi:10.20944/preprints202604.0365.v1



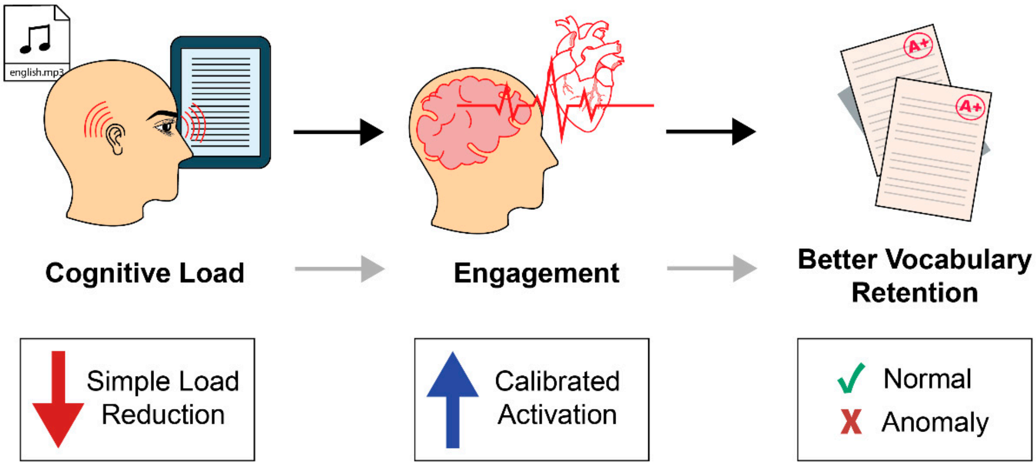


**Figure 2.** Balanced efficiency model of multimodal digital listening-while-reading proposed by this study.

## 7. Conclusion

The present study examined the role of multimodal listening-while-reading in digital EFL vocabulary learning by combining behavioral learning outcomes with physiological indicators of task engagement. The results provide evidence that synchronized auditory and textual input can significantly improve vocabulary acquisition compared to text-only reading. Learners in the listening-while-reading condition demonstrated substantial vocabulary gains, suggesting that coordinated multimodal input may strengthen lexical encoding and facilitate the integration of phonological and orthographic representations.

However, the physiological findings indicate that the mechanism underlying this improvement may be more complex than simple cognitive load reduction. Instead of showing lower physiological activation, learners in the listening-while-reading condition displayed moderately higher task-related heart rate levels. When considered together with the strong vocabulary outcomes and stable recovery patterns, this pattern appears more consistent with sustained attentional engagement than with cognitive overload.

Based on these findings, this study proposes a balanced efficiency model of multimodal digital listening-while-reading. According to this perspective, effective learning environments do not necessarily minimize cognitive effort. Rather, they regulate the relationship between cognitive demand and learner engagement. In listening-reading contexts, synchronized audio may reduce decoding-related barriers while simultaneously sustaining attentional focus on lexical processing. The resulting learning state is characterized not by minimal activation but by calibrated activation, in which cognitive resources are directed toward productive processing.

From a theoretical perspective, the findings suggest the following proposition: in second language vocabulary learning, optimal development may occur when cognitive load and learner engagement are jointly calibrated rather than minimized independently. Within this balanced efficiency framework, moderate physiological activation can reflect productive attentional mobilization that supports lexical integration.

More broadly, this study contributes to an emerging direction in SLA research that integrates behavioral outcomes with process-level indicators. By situating vocabulary gains within a cognitive-affective regulatory framework, the findings highlight the importance of examining not only whether learning improves, but under what activation profile it improves. Future research incorporating longitudinal modeling, additional physiological indices, and diverse learner populations will further clarify how calibrated activation supports sustainable language development.

In conclusion, the present findings suggest that listening-while-reading enhances vocabulary learning not by making processing uniformly easier, but by optimizing the balance between cognitive efficiency and sustained engagement. Within this biopsychological perspective, learning efficiency is not the absence of activation, but its effective regulation.

**Author Contributions:** Conceptualization, H.S.A. and C.L.; methodology, formal analysis, investigation, data curation, H.S.A.; writing—original draft preparation, H.S.A.; writing—review and editing, H.S.A. and C.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study because it involved voluntary participation and anonymous data collection from adult participants. Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

**Data Availability Statement:** The data presented in this study are available from the first author upon reasonable request.

**Acknowledgments:** Not applicable.:

**Conflicts of Interest:** The authors declare no conflict of interest.

## References


1. Nation, I.S.P.; Hunston, S. *Learning Vocabulary in Another Language*; 2nd ed.; Cambridge University Press: Cambridge, 2013; ISBN 9781107045477.
2. Schmitt, N. Understanding Vocabulary Acquisition, Instruction, and Assessment: A Research Agenda. *Language Teaching* **2019**, *52*, 261–274, doi:DOI: 10.1017/S0261444819000053.
3. Hulstijn, J.H. *Language Proficiency in Native and Non-Native Speakers*; John Benjamins Publishing Company: Amsterdam, 2015; Vol. 41; ISBN 9789027213242.
4. Webb, Stuart.; Nation, I.S.P.. *How Vocabulary Is Learned*; Oxford University Press, 2017; ISBN 9780194403559.
5. Ellis, N.C. At the Interface: Dynamic Interactions of Explicit and Implicit Language Knowledge. *Stud. Second Lang. Acquis.* **2005**, *27*, 305–352, doi:DOI: 10.1017/S027226310505014X.
6. Wulff, S.; Ellis, N.C. Usage-Based Approaches to Second Language Acquisition. In *Bilingual cognition and language: The state of the science across its subfields*; Miller, D., Bayram, F., Rothman, J., Serratrice, L., Eds.; John Benjamins Publishing Company: Amsterdam, 2018; Vol. 54, pp. 37–56.
7. Mayer, R.E.. *Multimedia Learning*; 3rd ed.; Cambridge University Press: Cambridge, 2021; ISBN 9781107187504.
8. Mayer, R.E.; Moreno, R. Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educ. Psychol.* **2003**, *38*, 43–52, doi:10.1207/S15326985EP3801_6.
9. Sweller, J. Cognitive Load During Problem Solving: Effects on Learning. *Cogn. Sci.* **1988**, *12*, 257–285, doi:https://doi.org/10.1207/s15516709cog1202_4.
10. Sweller, J.; Ayres, P.; Kalyuga, S. *Cognitive Load Theory*; Springer New York: New York, NY, 2011; ISBN 978-1-4419-8125-7.
11. Hiver, Phil.; Al-Hoorie, A.H..; Mercer, Sarah. *Student Engagement in the Language Classroom*; Multilingual Matters, 2021; ISBN 9781788923590.
12. Paas, F.; Ayres, P. Cognitive Load Theory: A Broader View on the Role of Memory in Learning and Education. *Educ. Psychol. Rev.* **2014**, *26*, 191–195, doi:10.1007/s10648-014-9263-5.
13. Pekrun, R. The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educ. Psychol. Rev.* **2006**, *18*, 315–341, doi:10.1007/s10648-006-9029-9.
14. Philp, J.; Duchesne, S. Exploring Engagement in Tasks in the Language Classroom. *Annu. Rev. Appl. Linguist.* **2016**, *36*, 50–72, doi:DOI: 10.1017/S0267190515000094.
15. Svalberg, A.M.-L. Engagement with Language: Interrogating a Construct. *Language Awareness* **2009**, *18*, 242–258, doi:10.1080/09658410903197264.
16. Shaffer, F.; Ginsberg, J.P. An Overview of Heart Rate Variability Metrics and Norms. *Front. Public Health* **2017**, *Volume 5-2017*.

17. Thayer, J.F.; Åhs, F.; Fredrikson, M.; Sollers, J.J.; Wager, T.D. A Meta-Analysis of Heart Rate Variability and Neuroimaging Studies: Implications for Heart Rate Variability as a Marker of Stress and Health. *Neurosci. Biobehav. Rev.* **2012**, *36*, 747–756, doi:https://doi.org/10.1016/j.neubiorev.2011.11.009.
18. Godfroid, A. *Eye Tracking in Second Language Acquisition and Bilingualism*; 1st ed.; Routledge: New York, 2019; ISBN 9781315775616.
19. Marsden, E.; Plonsky, L. Data, Open Science, and Methodological Reform in Second Language Acquisition Research. In *Critical Reflections on Data in Second Language Acquisition*; Edmonds, A., Gudmestad, A., Eds.; John Benjamins Publishing Company: Amsterdam, 2018; Vol. 51, pp. 219–228.
20. Plonsky, L. *Advancing Quantitative Methods in Second Language Research*; 1st ed.; Routledge: New York, 2015; ISBN 9781315870908.
21. Paivio, Allan. *Mental Representations : A Dual Coding Approach*; Oxford University Press: New York, 2010; ISBN 019503936X.
22. MacLeod, C.M.; Bodner, G.E. The Production Effect in Memory. *Curr. Dir. Psychol. Sci.* **2017**, *26*, 390–395, doi:10.1177/0963721417691356.
23. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*; 2nd ed.; Routledge: New York, 1988; ISBN 9781134742707.
24. Robinson, P. Attention and Memory during SLA. In *The Handbook of Second Language Acquisition*; 2003; pp. 631–678 ISBN 9780470756492.
25. Schumann, J.H.. *The Neurobiology of Affect in Language John H. Schumann*; Blackwell Publishing, 1999; ISBN 9780631210108.
26. Verspoor, M.; Lowie, W. Complex Dynamic Systems Theory and Second Language Development. In *Research Questions in Language Education and Applied Linguistics*; Mohebbi, H., Coombe, C., Eds.; 2021; pp. 799–803.
27. Plonsky, L.; Oswald, F.L. How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Lang. Learn.* **2014**, *64*, 878–912, doi:https://doi.org/10.1111/lang.12079.