From Exposure to Engagement: Subtitle-Mediated Vocabulary Learning in Secondary EFL Classrooms

Lan Chi Le Nguyen[1] and Gargi Sen[2]

[1,2] East Bridge University

Abstract: While previous research has demonstrated that subtitles can facilitate second language vocabulary learning, most studies rely on experimental designs that measure learning outcomes rather than examining how vocabulary engagement unfolds in authentic classroom contexts. Process-oriented classroom analyses of noticing and regulation during subtitle-supported lessons remain underrepresented.

This study investigates the cognitive and pedagogical processes underlying vocabulary engagement during English-subtitled video lessons in three Vietnamese upper-secondary EFL classrooms. Drawing on classroom observations, learner interviews, teacher interviews, and learner reflections, cross-case thematic analysis traced patterns of attention allocation, noticing, regulation, and mediation.

Findings indicate that vocabulary engagement did not occur automatically as a result of subtitle exposure. Instead, learners selectively attended to lexical items under multimodal constraints, and noticing was shaped by participation norms and teacher mediation. Across cases, engagement followed a recurring trajectory of selective noticing → verification/refinement → consolidation, with variation in regulation orientation (peer-supported, self-regulated, teacher-mediated).

The study reconceptualises subtitle-mediated vocabulary learning as a regulation-sensitive and pedagogically contingent classroom process, shifting attention from treatment effects to engagement mechanisms in instructional ecologies. Implications are discussed for multimodal pedagogy and process-oriented SLA research.

Keywords: subtitle-mediated vocabulary learning; noticing; learner regulation; classroom interaction; multimodal instruction; Vietnamese EFL.

1. Introduction

Audiovisual materials accompanied by subtitles have become increasingly common in English as a Foreign Language (EFL) classrooms. With the widespread availability of digital media and online video platforms, teachers frequently integrate English-subtitled videos to support listening comprehension, introduce thematic content, and expose learners to authentic language use. A growing body of research suggests that subtitles can facilitate vocabulary learning by enhancing the perceptual salience of lexical forms and strengthening form–meaning connections (Vanderplank, 2016; Peters & Webb, 2018; Montero Perez, 2022). In experimental and quasi-experimental studies, learners exposed to captioned or subtitled input often demonstrate higher vocabulary gains compared to those who view videos without textual support.

Despite these encouraging findings, much of the existing literature adopts outcome-oriented designs that measure vocabulary acquisition after exposure under controlled conditions. While such studies provide valuable evidence of facilitative effects, they offer limited insight into how vocabulary learning processes unfold in authentic classroom environments. In laboratory contexts,

learners typically watch videos individually, without peer interaction or teacher mediation, and vocabulary knowledge is assessed through post-tests. These conditions may not adequately capture the complex dynamics of real instructional settings, where attention is distributed across multiple modalities and shaped by classroom interaction.

From a cognitive perspective, vocabulary learning is closely linked to noticing, defined as the conscious registration of linguistic form in input (Schmidt, 1990, 2001). However, attention is inherently selective and constrained by limited processing capacity (Godfroid, 2016). In multimodal learning environments, learners must simultaneously manage auditory input, written subtitles, and visual information. According to cognitive theories of multimedia learning (Mayer, 2020), processing resources are finite, and learners must prioritise certain channels over others under time pressure. The presence of subtitles may increase lexical visibility, but visibility alone does not guarantee noticing or durable learning.

In addition to cognitive constraints, classroom learning is shaped by interactional and pedagogical factors. Ecological perspectives on language learning conceptualise classrooms as activity systems in which tools, participants, and tasks interact dynamically (van Lier, 2004). Within such systems, teachers structure opportunities for interpretation, clarification, and consolidation through instructional framing and mediation (Walsh, 2011). Peer interaction may also create spaces for collaborative meaning construction, influencing how learners attend to and process lexical items. Consequently, subtitle-mediated vocabulary learning cannot be fully understood without examining the interactional conditions under which it occurs.

Although previous research acknowledges the importance of attention and salience, relatively few studies have investigated how noticing and vocabulary engagement unfold during subtitle-supported lessons in naturalistic classroom contexts. The majority of studies treat subtitles as an input enhancement technique whose effects are measured quantitatively. Less attention has been paid to how learners allocate attention in real time, how they verify or refine emerging interpretations, and how teacher mediation shapes these processes. As a result, there remains a need for process-oriented research that traces vocabulary learning trajectories in authentic instructional settings.

The present study addresses this gap by examining vocabulary engagement during English-subtitled video lessons in three Vietnamese upper-secondary EFL classrooms. Rather than focusing on vocabulary gains as measurable outcomes, the study investigates how learners allocate attention, notice lexical items, and consolidate meaning within classroom interaction. By integrating cognitive perspectives on noticing with ecological and discourse-based views of classroom mediation, the study aims to provide a more nuanced understanding of subtitle-mediated vocabulary learning as a situated process.

The study is guided by the following research questions:

1. How do learners allocate attention and notice lexical items during English-subtitled video lessons in classroom settings?

2. How do teachers mediate learners' engagement with subtitled input across lesson stages?

3. How do cognitive and pedagogical processes interact to shape vocabulary learning trajectories in subtitle-supported instruction?

By foregrounding process rather than product, this study contributes to a growing body of research that seeks to understand how vocabulary learning emerges through the interplay of attention, interaction, and pedagogical design. In doing so, it offers insights relevant to both multimodal input research and classroom-based language teaching practice.

2. Literature Review

2.1 Subtitles and Vocabulary Learning in SLA Research

The use of subtitles and captions in second language (L2) learning has received sustained attention over the past two decades. A substantial body of empirical research suggests that captioned or subtitled audiovisual input can facilitate comprehension and support vocabulary acquisition (Vanderplank, 2016; Peters & Webb, 2018; Montero Perez, 2022). Subtitles are generally understood to enhance lexical salience by presenting written forms alongside auditory input, thereby strengthening form–meaning connections and increasing opportunities for noticing.

Experimental studies frequently report that learners exposed to L2 subtitles demonstrate higher vocabulary gains than those who view videos without textual support. Peters and Webb (2018), for example, found that captioned viewing promoted both incidental vocabulary learning and meaning recall. Similarly, Montero Perez (2022) synthesised evidence indicating that subtitles can positively influence vocabulary uptake, particularly when learners encounter repeated lexical items. These findings are often interpreted within an input-enhancement framework, which assumes that increasing perceptual salience heightens the likelihood of noticing and subsequent acquisition.

However, much of this research has been conducted under controlled laboratory conditions, where learners watch videos individually and vocabulary gains are assessed through immediate or delayed post-tests. While such designs offer strong internal validity, they may not capture the complexities of classroom learning. In instructional contexts, learners interact with peers, respond to teacher mediation, and allocate attention under time constraints. Consequently, subtitle effects observed in laboratory studies may not directly translate to authentic classroom settings.

Moreover, quantitative outcome measures reveal little about how learners process lexical input during viewing. Whether learners actively attend to unfamiliar words, how they interpret them, and how meanings become stabilised over time remain underexplored in many experimental designs. This suggests a need to complement outcome-based research with process-oriented classroom studies.

2.2 Noticing and Attentional Constraints

The theoretical foundation for much subtitle research lies in the noticing hypothesis (Schmidt, 1990, 2001), which posits that conscious attention to linguistic form is a necessary condition for acquisition. According to this view, learners must register lexical items in input before these items can be integrated into developing interlanguage systems. Subtitles, by rendering lexical forms visually explicit, are therefore assumed to increase the likelihood of noticing.

However, attention is limited and selective. Cognitive research in SLA demonstrates that learners cannot attend to all aspects of input simultaneously (Godfroid, 2016). Instead, attention is distributed according to perceived relevance, task demands, and processing capacity. In multimodal environments, learners must manage written text, auditory speech, and visual imagery concurrently. This raises the possibility that subtitles may compete with or redirect attention rather than uniformly enhance it.

Recent work in eye-tracking and attentional research further indicates that noticing is gradient rather than binary (Godfroid et al., 2018). Learners may partially register lexical items without fully processing their meanings. Thus, while subtitles increase exposure to orthographic forms, exposure alone does not guarantee deep processing or durable learning. These insights challenge simplistic interpretations of subtitles as automatic facilitators of noticing and highlight the importance of examining how attention operates in real-time instructional contexts.

2.3 Multimedia Processing and Cognitive Load

Cognitive theories of multimedia learning (Mayer, 2020) provide additional insight into subtitle-mediated learning. According to dual-channel processing models, learners process auditory and visual information through partially separate channels, both of which have limited capacity. When learners view subtitled videos, they must allocate cognitive resources to reading text while simultaneously listening and attending to visual cues. Under time pressure, learners may prioritise one channel over another.

Research suggests that subtitles can reduce listening difficulty by supporting comprehension; however, they may also increase cognitive load if learners struggle to synchronise reading and listening. Split-attention effects may occur when learners divide their focus between competing sources of information. Therefore, the effectiveness of subtitles depends not only on their presence but also on how learners regulate attention across modalities.

While multimedia research has largely focused on cognitive load and comprehension outcomes, fewer studies have examined how these processing dynamics interact with classroom participation and pedagogy. Understanding subtitle-mediated vocabulary learning requires attention to both cognitive constraints and interactional organisation.

2.4 Classroom Mediation and Ecological Perspectives

From an ecological perspective, language learning is understood as emerging through participation in socially organised activity systems (van Lier, 2004). Classrooms are not neutral environments; they are structured by interactional norms, teacher mediation, and task design. Teachers influence learning opportunities by framing tasks, pausing input, prompting explanation, and orchestrating discussion (Walsh, 2011). Such mediation may shape how learners attend to and interpret lexical items during subtitle-supported lessons.

Peer interaction also plays a role in distributed cognition. Collaborative dialogue can facilitate meaning construction and provide opportunities for verification and clarification. In this sense, noticing may not be solely an individual cognitive event but an interactionally mediated process.

Despite recognition of these classroom dynamics, relatively few studies have investigated subtitle-mediated vocabulary learning through qualitative classroom analysis. Existing research tends to treat subtitles as a technical affordance whose effects are measurable, rather than as a resource embedded within pedagogical and interactional structures.

2.5 Research Gap

Taken together, prior research establishes that subtitles can support vocabulary learning, but it leaves underexplored how noticing and consolidation unfold in authentic classroom environments. Laboratory findings demonstrate facilitative effects, yet they provide limited insight into attentional management, regulation strategies, and teacher mediation during real lessons.

While caption research has expanded significantly, process-oriented classroom analyses of noticing and regulation during subtitle-supported lessons remain underrepresented. Most existing studies conceptualise subtitles as input enhancement variables and assess their effects through post-test measures, offering limited insight into how vocabulary engagement unfolds in authentic instructional settings.

There is therefore a need for classroom-based research that examines subtitle-mediated vocabulary engagement as a situated phenomenon shaped by cognitive constraints and pedagogical organisation. By tracing attention allocation, noticing trajectories, and consolidation practices in naturalistic lessons, such research can theorise how subtitles function as learning resources in practice rather than as isolated treatment variables.

3. Methodology

3.1 Research Design

This study adopted a qualitative multiple-case design to investigate how vocabulary learning processes unfold during English-subtitled video lessons in authentic classroom contexts. A qualitative approach was chosen because the research questions focus on tracing attention allocation, noticing trajectories, and pedagogical mediation as they occur in real time, rather than measuring vocabulary gains quantitatively. The study therefore prioritised process-oriented analysis over outcome-based evaluation.

The multiple-case design enabled cross-case comparison across three intact classrooms while maintaining ecological validity. Treating each classroom as a bounded case allowed for in-depth examination of interactional patterns within each context, followed by systematic comparison to identify recurring mechanisms and contextual variation.

3.2 Research Context and Participants

The study was conducted in a public upper-secondary school in Vietnam where English is taught as a compulsory subject within the national curriculum. English instruction emphasises vocabulary knowledge as a foundation for reading comprehension and examination preparation. Class sizes are relatively large, and lessons are typically teacher-led, although multimedia materials are increasingly incorporated.

Three Grade 11 classes (11A6, 11A7, and 11A15) participated in the study. Approximately 150 students were enrolled across the three classes. All classes were taught by the same experienced English teacher with more than ten years of teaching experience. The teacher regularly integrated English-subtitled videos into lessons to support comprehension and vocabulary learning.

Using the same teacher across cases allowed the study to examine how subtitle-supported instruction unfolded under relatively stable pedagogical conditions while exploring variation in classroom participation and learner engagement.

Participation in interviews and reflection activities was voluntary. Informed consent was obtained from students and guardians, and pseudonyms were used to protect anonymity.

3.3 Data Collection

Data were collected across six subtitle-supported lessons (approximately five hours of observation in total). Each observed lesson included pre-viewing, viewing, and post-viewing stages.

Multiple qualitative data sources were employed to capture both observable classroom interaction and participants' perspectives:

Classroom Observations
Lessons were observed and documented through detailed field notes. Observation focused on:

- Learners' visible attention allocation (e.g., gaze orientation, note-taking, underlining)
- Peer interaction during viewing
- Teacher mediation practices (pausing, prompting, clarification)
- Post-viewing consolidation tasks

Observation notes were structured according to lesson stages to facilitate later comparison.

Learner Interviews (n = 9)
Semi-structured interviews were conducted with selected students from the three classes. Interview questions explored:

- How learners allocated attention during video viewing
- What triggered noticing of unfamiliar words
- Strategies used to verify or remember vocabulary
- Perceived benefits and challenges of subtitles

Interviews were conducted in Vietnamese to ensure comfort and depth of response. Audio recordings were transcribed and translated into English for analysis.

Teacher Interview (n = 1)
A semi-structured interview with the teacher examined:

- Rationale for using English-subtitled videos
- Decisions regarding pausing and task design
- Perceptions of learners' engagement and difficulties

Learner Reflection Forms (n = 60)
Short open-ended reflection forms were administered after selected lessons. Students described:

- How they processed subtitles
- Whether and how they noticed new vocabulary

o   What helped them remember words

Reflexive Field Notes
The researcher maintained analytic memos documenting emerging interpretations and methodological decisions.

Data collection was iterative, with preliminary analysis informing subsequent observation focus.

Table 1 provides an overview of the dataset across the three focal classrooms.

Table 1. Overview of the dataset

Data source	Description	Quantity
School	Public upper-secondary school	1
Classes	Grade 11 (11A6, 11A7, 11A15)	3
Students	Approximate total enrolment	150
Lessons observed	Subtitle-supported lessons	6
Observation hours	Total classroom observation	~5 hours
Learner interviews	Semi-structured	9
Teacher interviews	Semi-structured	1
Reflection forms	Open-ended responses	60
Total transcript length	Transcribed data corpus	~80 pages

3.4 Data Analysis

Data were analysed using thematic analysis (Braun & Clarke, 2006). The analytic process involved six stages:

Familiarisation
Transcripts and field notes were read repeatedly to identify preliminary patterns related to attention, noticing, and mediation.

Initial Coding
Segments of data were coded inductively and deductively. Initial codes included:

o   Subtitle-first attention

o   Repetition-triggered noticing

o   Peer-assisted verification

o   Teacher-directed focus

o   Post-viewing articulation

Category Development
Codes were grouped into broader categories representing cognitive and pedagogical processes.

Theme Construction

Recurring patterns across cases were synthesised into higher-level themes describing vocabulary engagement trajectories.

Cross-Case Comparison

Patterns were compared across the three classrooms to identify similarities and contextual differences.

Refinement and Verification

Themes were refined through constant comparison across data sources (observations, interviews, reflections). Evidence extracts were revisited to ensure alignment between interpretation and data.

Rather than aiming for statistical representativeness, the analysis sought analytic saturation, defined as the point at which additional data confirmed rather than substantially altered emerging thematic patterns.

3.5 Trustworthiness and Research Quality

To enhance credibility, methodological triangulation was employed by integrating observation data, learner interviews, teacher interviews, and reflection forms. Convergence of evidence across sources strengthened interpretive claims.

An audit trail documenting coding decisions and theme development was maintained. Reflexive journaling was used to monitor potential researcher bias, particularly given the researcher's insider role within the school context.

Detailed contextual description supports analytic transferability to comparable EFL classroom settings.

4. Results

This section presents findings from cross-case thematic analysis of classroom observations, learner interviews, teacher interviews, and reflection data. Three major patterns emerged across the three classrooms: (1) selective attention and noticing under subtitle conditions, (2) verification and refinement through interaction and mediation, and (3) consolidation through articulation and task engagement. Although patterns were broadly comparable, variation across classrooms reflected differences in regulation orientation and mediation style.

4.1 Selective Attention and Noticing Under Subtitle Conditions

Across all three classrooms, learners did not attend uniformly to lexical items presented in subtitles. Instead, noticing was selective and shaped by time pressure, perceived relevance, and interactional cues.

During the viewing stage, most learners prioritised reading subtitles as the primary channel for comprehension. One learner explained: "I read the subtitle first, then I listen to check if the speaker says the same thing." (IntL-11A7-Lr09-Trans). Observation notes indicated sustained gaze orientation toward subtitle lines, often with minimal attention to visual imagery. During viewing, several students kept their eyes on the subtitle line and only looked up at the image after finishing each line. (Obs-11A15-L2-V-09:10–09:20). However, the degree and nature of this prioritisation varied across classes.

In Class 11A6, learners alternated between subtitle text and on-screen visuals while occasionally engaging in brief peer whisper-discussions. For example, when encountering unfamiliar vocabulary such as "evacuation," two students briefly negotiated meaning in Vietnamese before returning attention to the video.

One student whispered, "Is that like emergency?" before returning to the screen. (Obs-11A6-L3-V-Trans)

These exchanges typically lasted only two to three seconds and did not interrupt viewing flow. Such episodes suggest that noticing was distributed across peers, with subtitles serving as an initial trigger for collaborative meaning exploration.

In contrast, learners in Class 11A7 adopted a subtitle-first monitoring routine characterised by minimal overt peer interaction during viewing. Observation records indicated that students remained largely silent while reading subtitles line by line. Interview data confirmed that learners intentionally avoided talking during viewing to prevent missing subsequent subtitle lines. As one

learner explained, maintaining reading continuity was prioritised over immediate discussion. In this context, noticing occurred primarily as an individual, internally regulated process.

In Class 11A15, noticing was more explicitly shaped by teacher direction. During pre-viewing framing, the teacher instructed students to underline unfamiliar words and highlighted specific lexical targets. When these words appeared in subtitles, learners were observed underlining or quietly repeating them. This suggests that teacher mediation structured attentional focus and increased the likelihood of noticing particular items.

Across cases, repetition emerged as a salient noticing trigger. As one student commented, "When it appears again, I realise it must be important." (Ref-11A6-L5-ItemD4-Trans). Learners reported that when lexical items reappeared across scenes, they became more noticeable and more likely to be recorded. However, not all unfamiliar words were noticed. Attention appeared to be selective and influenced by perceived importance or task relevance.

Taken together, these findings indicate that subtitles increased lexical visibility but did not guarantee comprehensive noticing. Instead, noticing operated as a selective process shaped by attentional management and classroom conditions.

4.2 Verification and Refinement Through Interaction and Mediation

Noticing alone did not constitute vocabulary learning. Across classrooms, learners engaged in processes of verification and refinement to stabilise emerging interpretations.

In Class 11A6, verification frequently occurred through peer interaction. After initial noticing triggered by subtitles, learners briefly consulted partners to confirm meaning. These exchanges were often rapid and embedded within ongoing viewing. Learners used subtitles as reference points, sometimes comparing written forms with auditory input or contextual clues. Peer discussion functioned as a collaborative verification mechanism that strengthened tentative interpretations.

S6: "Look, it appears again."
S7: "Yeah, underline it."
(Obs-11A6-L4-V-09:22–09:27)

In Class 11A7, verification was more internally regulated. Learners described forming initial hypotheses about meaning from context and then using subtitles to confirm or revise those interpretations. For example, some learners reported that they first inferred meaning from visual cues or prior knowledge and subsequently checked alignment with subtitle text. When discrepancies emerged, learners adjusted their understanding. In this classroom, subtitles functioned primarily as confirmation tools rather than primary sources of meaning.

"When I don't catch the spoken English, I rely on the subtitles to understand the meaning." (IntL-11A7-Lr02-Trans)

In Class 11A15, verification was often mediated by the teacher. During post-viewing discussion, the teacher prompted students to explain or justify interpretations of targeted vocabulary.

Teacher: "Why do you think it means that? What in the video shows that?" (Obs-11A15-L5-PV-Trans).

When learners provided tentative responses, the teacher either expanded or clarified meanings. This dialogic mediation extended the processing phase beyond initial noticing and required learners to articulate understanding publicly.

Across cases, verification processes varied in degree of interactional visibility. In 11A6, they were peer-mediated and immediately enacted; in 11A7, they were internally regulated and sometimes delayed; in 11A15, they were teacher-orchestrated. Despite these differences, a common pattern was evident: learners did not treat subtitles as final authorities but as resources to confirm or refine evolving interpretations.

4.3 Consolidation Through Articulation and Task Engagement

The third recurring pattern involved consolidation of lexical items beyond the moment of noticing. Consolidation typically occurred during post-viewing stages and was closely linked to task design.

In Class 11A6, consolidation frequently involved rewriting and annotating vocabulary in notebooks. After viewing, learners were observed adding Vietnamese meanings next to underlined English words and occasionally constructing short example sentences. This suggests that articulation through writing supported stabilisation of form–meaning connections.

Several students were observed adding Vietnamese glosses next to underlined words immediately after viewing. (Obs-11A6-L5-PV-09:02–09:06)

In Class 11A7, consolidation was often delayed until after the video ended. Learners reported writing down difficult words and seeking clarification from peers or dictionaries.

"After the video, I write the new words in my notebook and check later." (IntL-11A7-Lr04-Trans)

"If I don't write it, I forget it quickly." (Ref-11A7-L5-ItemC2-Trans)

Task requirements, such as summarising video content or answering comprehension questions, increased the likelihood that new vocabulary would be revisited. Thus, consolidation was linked to subsequent task engagement rather than immediate discussion during viewing.

In Class 11A15, consolidation was strongly structured by teacher prompts. Post-viewing activities required students to use specific lexical items in sentences or explanations. When learners attempted to produce examples, the teacher provided corrective feedback and encouraged elaboration. This interaction transformed noticed words into shared objects of analysis and use.

When prompted to use the word in a sentence, one learner hesitated and then reformulated: *"Evacuation… means people must leave quickly because of danger." (Obs-11A15-L5-PV-Trans)*

Across classrooms, consolidation appeared strongest when learners were required to articulate or apply vocabulary rather than merely recognise it. Mere exposure or silent reading did not consistently lead to extended processing. Instead, tasks that prompted explanation, paraphrasing, or contextual reuse created opportunities for deeper engagement.

4.4 Cross-Case Synthesis: A Process Trajectory

Despite variation in participation norms and mediation styles, a recurring engagement trajectory was observable across the three classrooms. Vocabulary engagement unfolded through:

- Selective noticing triggered by subtitle salience or repetition
- Verification/refinement through peer interaction, internal monitoring, or teacher mediation
- Consolidation through articulation, recording, or task-based reuse

While the enactment of each phase differed, the underlying sequence remained stable. In 11A6, noticing and refinement were frequently distributed across peer interaction during viewing. In 11A7, learners tended to regulate processing internally during viewing and externalised interpretations later. In 11A15, teacher prompts structured both attentional focus and consolidation sequences.

Across cases, lexical items that were publicly articulated during post-viewing tasks were observed to reappear in subsequent turns of classroom talk, suggesting that articulation increased the durability of lexical engagement (Obs comparison notes). Lexical items that were not publicly articulated were less likely to reappear in subsequent interactional turns. As one learner reflected, "First I notice it, then I check if I'm right, and later I try to remember it." (IntL-11A7-Lr03-Trans).

Taken together, these cross-case patterns indicate that vocabulary engagement in subtitle-supported lessons was temporally distributed and regulation-dependent rather than immediate or uniform.

4.5 Summary of Findings

The findings indicate that subtitle-mediated vocabulary engagement in classroom contexts was characterised by selective attention under multimodal constraints, regulation-dependent verification processes, and consolidation supported by articulation and task enactment.

Across cases, subtitles functioned as visible lexical resources; however, their learning value depended on how attention was organised and how opportunities for verification and reuse were structured within classroom activity. Vocabulary engagement did not emerge as an automatic consequence of exposure, but as a temporally distributed process shaped by interactional and pedagogical conditions.

5. Discussion

The findings of this study suggest that subtitle-mediated vocabulary learning in classroom contexts operates through regulation-sensitive engagement processes rather than automatic lexical uptake. Importantly, the results do not contradict experimental evidence showing that subtitles can facilitate vocabulary acquisition (Peters & Webb, 2018; Montero Perez, 2022). Instead, they refine the explanatory mechanism underlying such effects by demonstrating how engagement unfolds under real-time classroom constraints. Across the three cases, vocabulary engagement followed a recurring trajectory of selective noticing → verification/refinement → consolidation. However, the enactment of this trajectory depended on regulation orientation and mediation practices, indicating that subtitles operate within broader classroom ecologies rather than as independent causal variables.

5.1 Selective Noticing Under Multimodal Constraints

The findings confirm that subtitles increase lexical visibility, but they complicate the assumption that increased visibility ensures comprehensive noticing. Across cases, learners attended selectively to lexical items, prioritising subtitle reading as a strategy for maintaining comprehension continuity. This selective attention aligns with attention-based SLA accounts, which emphasise that noticing is constrained by limited processing capacity (Schmidt, 1990, 2001; Godfroid, 2016). In multimodal environments, learners must distribute attention across written, auditory, and visual channels. Under time pressure, prioritisation becomes inevitable.

What the present findings add is classroom-level specificity. Selectivity was not solely cognitive; it was shaped by participation norms and instructional framing. In one classroom, peer interaction redistributed attention and made noticing more interactionally visible. In another, silence functioned as a self-regulation strategy to preserve subtitle flow. In the third, teacher prompts pre-structured attentional focus. These differences indicate that noticing is not simply triggered by salience but organised through classroom participation structures.

In contrast to laboratory-based studies where exposure conditions are controlled and interaction is minimised, the present findings demonstrate that noticing in classrooms is interactionally organised and socially distributed.

5.2 Verification as a Regulation-Dependent Process

A central contribution of the study lies in demonstrating that noticing did not immediately stabilise lexical understanding. Instead, learners engaged in verification and refinement processes before meanings became more secure. This aligns with research on incremental vocabulary acquisition (Nation, 2013; Schmitt, 2010), which emphasises gradual strengthening of form–meaning connections.

However, the present findings move beyond incrementalism by specifying how verification is organised. In peer-mediated contexts, refinement occurred through distributed dialogue. In self-regulated contexts, learners internally compared contextual inference with subtitle confirmation. In teacher-mediated contexts, clarification was structured through prompt-and-response sequences. Although the surface interaction differed, the functional role of verification remained consistent: tentative interpretations were tested, confirmed, or revised.

Crucially, subtitles functioned primarily as monitoring tools rather than direct meaning transmitters. Learners frequently generated provisional hypotheses from contextual or visual cues and used subtitles to check alignment. This observation repositions subtitles from primary semantic sources to regulatory resources embedded in meaning-making trajectories. Rather than "providing vocabulary," subtitles supported learners' self- and socially regulated interpretive work.

From a sociocultural perspective (Lantolf & Thorne, 2006), this suggests that cognitive processes of noticing and refinement are mediated through available semiotic tools and participation frameworks. Verification is therefore neither purely internal nor purely instructional; it is distributed across learners, tools, and pedagogical structures.

5.3 Consolidation Through Articulation and Task Enactment

The final stage of the observed trajectory involved consolidation through articulation. Vocabulary engagement intensified when learners were required to explain, rewrite, or reuse lexical items in post-viewing tasks. This finding supports research on generative use and retrieval practice, which highlights the importance of productive engagement for retention (Nation, 2013).

More significantly, consolidation was contingent on task design. Where tasks required public articulation or example generation, lexical processing extended beyond recognition. Where tasks prioritised comprehension only, noticing sometimes remained partial. This indicates that subtitles alone do not produce durable learning; consolidation requires pedagogically structured opportunities for transformation and reuse.

The findings thus support a pedagogical orchestration view (Walsh, 2011), in which teachers shape processing depth through timing, prompting, and task sequencing. Subtitles create potential affordances, but these affordances become learning opportunities only when activated through classroom activity.

5.4 Reconceptualising Subtitle-Mediated Vocabulary Learning

Taken together, the findings support a reconceptualisation of subtitle-mediated vocabulary learning as a regulation-sensitive and pedagogically contingent process. Unlike experimental studies that conceptualise subtitles primarily as input-enhancement variables, the present analysis positions subtitles as interactionally embedded resources whose learning value depends on attention management and mediation practices.

The recurring trajectory observed across cases—selective noticing → verification/refinement → consolidation—should not be interpreted as a fixed instructional sequence. Rather, it represents a process model describing how vocabulary engagement becomes learnable within classroom ecologies. Noticing is not a discrete event triggered by salience; it is part of a temporally distributed engagement pathway shaped by regulation orientation (shared, self-regulated, teacher-mediated).

This perspective shifts the analytical focus from treatment effects to engagement mechanisms. By reframing subtitles as interactionally embedded resources rather than input-enhancement variables, the study invites a reconsideration of how multimodal tools are theorised in classroom SLA research. It challenges outcome-dominant paradigms in caption research by arguing that subtitle effectiveness cannot be reduced to exposure conditions alone. Instead, subtitles function within systems of participation, mediation, and task enactment. Their impact emerges through how learners and teachers organise attention and interaction around them.

5.5 Implications for Multimodal SLA Research

The study contributes theoretically by bridging attention-based SLA frameworks with ecological classroom perspectives. Attention remains a necessary condition for learning, but it is insufficient without regulatory support and consolidation opportunities. Multimodal tools such as subtitles must therefore be analysed not only in terms of perceptual salience but also in terms of how they are integrated into instructional activity systems.

By foregrounding process rather than product, this research invites SLA scholarship to examine engagement trajectories alongside measurable gains. Classroom-based studies can complement laboratory experiments by revealing how learning mechanisms operate under authentic constraints.

Pedagogically, the findings suggests that effective subtitle integration requires deliberate orchestration of attention, timing of mediation, and structured opportunities for articulation, rather than reliance on textual support alone.

5.6 Limitations and Future Directions

As a qualitative multiple-case study conducted within one institutional context, the findings prioritise analytic depth over statistical generalisation. The study traced engagement processes rather than long-term vocabulary retention. Future research combining classroom interaction analysis with delayed outcome measures could clarify how observed engagement trajectories relate to durable acquisition.

Comparative investigations across different subtitle conditions or instructional designs may further illuminate how regulation orientations shift under varying multimodal configurations.

5.7 Discussion Summary

Overall, the findings demonstrate that subtitle-mediated vocabulary learning is not an automatic outcome of enhanced lexical visibility. Instead, it unfolds as a selective, regulation-dependent, and pedagogically organised process. Subtitles provide lexical affordances, but their effectiveness depends on how attention is managed, how meanings are verified, and how consolidation is enacted within classroom ecologies.

6. Conclusion

This study examined how vocabulary engagement unfolds during English-subtitled video lessons in authentic secondary EFL classrooms. Rather than treating subtitles as input-enhancement devices whose effects can be evaluated solely through outcome measures, the analysis foregrounded the processes through which attention, verification, and consolidation were organised within classroom interaction.

The findings indicate that subtitle-mediated vocabulary learning cannot be reduced to lexical visibility or exposure alone. Instead, engagement emerged as a regulation-sensitive and pedagogically contingent process shaped by participation norms, mediation practices, and task enactment. Subtitles functioned not as self-sufficient scaffolds, but as semiotic resources whose learning value depended on how learners and teachers structured attention and opportunities for articulation.

By shifting the analytical focus from treatment effects to engagement mechanisms, the study contributes a classroom-grounded perspective to research on multimodal vocabulary learning. It demonstrates that understanding subtitle effectiveness requires examining how multimodal tools are enacted within instructional ecologies rather than isolating them as independent variables.

The study also highlights the importance of pedagogical orchestration. Effective subtitle integration involves more than providing textual support; it requires deliberate management of attention, structured moments for verification, and tasks that transform initial noticing into articulated understanding. These findings suggest that multimodal affordances become pedagogically meaningful only when activated through interaction.

While limited to three classrooms within one institutional context, the research offers analytically transferable insights into how vocabulary engagement unfolds under authentic constraints. Future work integrating process-oriented classroom analysis with longitudinal outcome measures may further illuminate how regulation-dependent engagement trajectories relate to durable lexical development.

Disclosure Statement
The author declares that there are no conflicts of interest.

References

[1]. **Braun, V., & Clarke, V. (2006).** Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

[2]. **Godfroid, A., Boers, F., & Housen, A. (2018).** An eye for words: Gauging the role of attention in incidental L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition, 40*(1), 1–36. https://doi.org/10.1017/S0272263117000109

[3]. **Lantolf, J. P., & Thorne, S. L. (2006).** *Sociocultural theory and the genesis of second language development*. Oxford University Press.

[4]. **Mayer, R. E. (2020).** *Multimedia learning* (3rd ed.). Cambridge University Press.

[5]. **Nation, I. S. P. (2013).** *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

[6]. **Peters, E., & Webb, S. (2018).** [verify details: exact title/journal/pages/DOI].

[7]. **Schmidt, R. (1990).** The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.

[8]. **Schmidt, R. (2001).** Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

[9]. **Schmitt, N. (2010).** *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

[10]. **van Lier, L. (2004).** *The ecology and semiotics of language learning: A sociocultural perspective*. Springer.

[11]. **Vanderplank, R. (2016).** *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. Palgrave Macmillan.

[12]. **Walsh, S. (2011).** *Exploring classroom discourse: Language in action*. Routledge.

[13]. **Montero Perez, M. (2022).** [verify details: exact source you mean in 2022—book/chapter/article title, venue, pages, DOI].