

FOREIGN LANGUAGES AND VIETNAMESE

MULTIMODAL VOCABULARY ACQUISITION IN AVIATION ENGLISH: LEARNER PERCEPTIONS THROUGH THE LENS OF MLT, DCT, AND CLT**NGUYEN THI NHU DIEP***

ABSTRACT: This study examines students' perceptions of multimodal vocabulary learning in Aviation English, drawing on Multimedia Learning Theory (MLT), Dual Coding Theory (DCT), and Cognitive Load Theory (CLT). A survey of 266 undergraduates at the Vietnam Aviation Academy revealed a strong preference for multimodal, context-based approaches. Learners considered visual input, real-life or simulated scenarios, and contextual analysis the most effective means of supporting vocabulary retention and retrieval, whereas written text alone was regarded as insufficient. More than 87% of participants reported difficulties in memorizing and visualizing terms when relying solely on print materials. The findings highlight the linguistic value of integrating verbal, visual, and contextual modalities, demonstrating how multimodal input enhances the acquisition, storage, and use of specialized vocabulary in professional communication.

KEY WORDS: Aviation English; Multimedia Learning Theory; Dual Coding Theory; Visual input; Contextual analysis.

RECEIVED: September, 30th, 2025.

EDITED: November, 15th, 2025

1. Introduction

Vocabulary constitutes a critical component of Aviation English, a branch of English for Specific Purposes (ESP) where accuracy and clarity in communication are paramount under high-stakes conditions. Effective vocabulary acquisition in this domain requires more than rote memorization of terms; it involves contextual understanding, rapid retrieval, and the ability to apply lexical knowledge in both spoken and written communication. While traditional vocabulary instruction in Vietnam has often relied on text-based memorization, such approaches have proven insufficient in addressing the cognitive and communicative demands of aviation contexts.

In recent years, multimodal approaches—combining text, images, audio, and video—have been increasingly recognized as promising strategies for vocabulary learning. International research has provided strong empirical evidence for their effectiveness. Li et al. (2022) demonstrated that multimodal input enhances short-term retention but must be carefully designed to avoid cognitive overload. Huang et al. (2022) found that video-based multimodal learning significantly improved learners' acquisition of word phrases compared to traditional methods. Similarly, Godfroid et al. (2017) used eye-tracking data to show that visual support increases learner attention to unfamiliar words. These findings support the growing consensus that multimodal learning not only enhances memory but also boosts learner engagement and motivation.

Despite these international advances, research in the Vietnamese context has been relatively limited and largely material oriented. Nguyen and Tran (2023), for instance, developed a bilingual multimodal electronic dictionary that improved student motivation and retention, though its scope remained restricted to vocabulary lists rather than communicative practice. Pham and Le (2022)

*Vietnam Aviation Academy; Email: diepntn@vaa.edu.vn

conducted a needs analysis among air traffic control students, revealing persistent difficulties in applying vocabulary in listening and speaking, and highlighting the urgent demand for more contextualized ESP curricula. Likewise, Do, Nguyen, and Huynh (2021) evaluated an ICAO-oriented textbook, recognizing its relevance but noting a lack of interactive vocabulary practice. Collectively, these studies underline that vocabulary remains the principal barrier for Vietnamese aviation students, while also revealing a research gap: the absence of systematic investigations into multimodal vocabulary learning informed by cognitive theory.

To address this gap, the present study is guided by three key theoretical perspectives. The Multimedia Learning Theory (MLT; Mayer, 2009) posits that learning is more effective when verbal and visual inputs are combined. The Dual Coding Theory (DCT; Paivio, 1986) emphasizes that information encoded through both linguistic and non-linguistic systems produces richer memory traces and facilitates recall. Meanwhile, the Cognitive Load Theory (CLT; Sweller, 1998) stresses the importance of instructional design in preventing working memory overload when integrating multiple input modes. Together, these frameworks suggest that multimodal vocabulary learning has the potential to enhance retention and motivation, provided that it is implemented with careful attention to cognitive constraints.

Building on established theoretical foundations, this study investigates Vietnamese learners' perceptions of multimodal strategies for acquiring Aviation English vocabulary. It is guided by two central research questions: 1/How do aviation English learners in Vietnam perceive and prioritize different multimodal techniques for vocabulary development?; 2/To what extent do these perspectives converge or diverge, and what linguistic insights can be drawn for vocabulary learning in ESP contexts?

By addressing these questions, the study aims to advance theoretical understanding of multimodal input in vocabulary acquisition while also illustrating its potential applications to the teaching and learning of Aviation English in Vietnamese higher education.

2. Literature Review

2.1. International Research on Multimodal Vocabulary Learning

Research on multimodal vocabulary instruction has expanded significantly in recent years, drawing on cognitive frameworks such as Multimedia Learning Theory (Mayer, 2009), Dual Coding Theory (Paivio, 1986), and Cognitive Load Theory (Sweller, 1998). Numerous empirical studies provide evidence of the effectiveness of multimodal approaches.

For instance, Li et al. (2022) compared learners exposed to text-only input with those receiving combined textual, visual, and auditory input. Their findings indicated that the multimodal group achieved higher short-term retention, although excessive input sometimes resulted in cognitive overload when instructional design was not well structured. Huang et al. (2022) demonstrated that integrating video, audio, and note-taking strategies significantly enhanced the acquisition of multiword units, suggesting that multimodal learning supports not only single-word recall but also phrase-level competence. Eye-tracking research by Godfroid et al. (2017) showed that learners allocated greater attention to unfamiliar vocabulary when supported by images or glosses, reinforcing the idea that multimodal scaffolding increases attentional focus.

Collectively, these studies suggest that multimodal vocabulary learning enhances learner motivation, deepens processing, and strengthens memory traces. However, they also point to the importance of instructional design, particularly in balancing input complexity to prevent overload.

2.2. Vietnamese Research on Aviation English Vocabulary Learning

In Vietnam, research into Aviation English vocabulary instruction has been relatively limited and tends to focus on teaching materials and learner needs rather than cognitive frameworks.

Nguyen and Tran (2023) developed and evaluated a bilingual multimodal electronic dictionary using Book Creator. Their findings showed that students demonstrated higher motivation and better

vocabulary retention when using the tool, although the study primarily focused on vocabulary lists rather than communicative application. Pham and Le (2022) conducted a needs analysis of air traffic control students, revealing that learners struggled to apply vocabulary in listening and speaking tasks, and recommending that future ESP syllabi incorporate more communicative activities. Similarly, Do, Nguyen, and Huynh (2021) evaluated the textbook *Aviation English for ICAO Compliance*, concluding that while the material was relevant to ICAO standards, it lacked sufficient interactive tasks to support active vocabulary use.

Taken together, these studies underscore that vocabulary remains a major obstacle for Vietnamese aviation students. At the same time, they highlight an urgent demand for more contextually grounded and communicatively oriented ESP curricula.

2.3. Research Gap

While international studies provide strong evidence that multimodal approaches can improve vocabulary learning, Vietnamese research in the field of Aviation English remains underdeveloped. Current studies primarily emphasize the evaluation of teaching materials and learner needs, rather than systematically testing the cognitive and pedagogical benefits of multimodal vocabulary instruction.

This gap suggests the need for empirical studies that integrate global theoretical frameworks (MLT, DCT, CLT) with the Vietnamese educational context, focusing specifically on students' perspectives as an initial step. While the present study provides insights into learners' preferences, future research should incorporate teachers' viewpoints to gain a more comprehensive understanding of how multimodal approaches can be effectively adapted to aviation-related ESP programs and aligned with instructional design and institutional realities.

3. Theoretical Framework

3.1. Multimedia Learning Theory (MLT)

The Multimedia Learning Theory (Mayer, 2009) posits that learners acquire knowledge more effectively when information is presented through multiple modes—typically verbal (spoken or written text) and visual (images, graphics, or video). The theory is based on three core assumptions: (1) humans have dual channels for processing verbal and visual information; (2) each channel has limited capacity; and (3) meaningful learning occurs when learners actively select, organize, and integrate information. In the context of Aviation English, where terminology often requires both precise definitions and contextualized visual representation (e.g., cockpit instruments, airport signage), MLT provides a strong rationale for incorporating multimodal materials into vocabulary instruction.

3.2. Dual Coding Theory (DCT)

The Dual Coding Theory (Paivio, 1986) complements MLT by emphasizing that information is encoded through two interconnected systems: the verbal system (language-based) and the non-verbal system (imagery-based). When both systems are activated simultaneously—for instance, when learners encounter a new aviation term with a definition, pronunciation, and a supporting diagram—they develop richer memory traces. This dual representation not only facilitates recall but also enhances transfer, as learners can retrieve the concept through either linguistic or visual pathways. For specialized vocabulary in aviation, DCT suggests that combining technical definitions with authentic visuals (e.g., radar screens, flight procedures) strengthens long-term retention and practical application.

3.3. Cognitive Load Theory (CLT)

While MLT and DCT highlight the benefits of multimodal input, the Cognitive Load Theory (Sweller, 1998) cautions against overloading learners' working memory. CLT distinguishes between three types of cognitive load: intrinsic load (complexity of the material), extraneous load (unnecessary information or poor design), and germane load (cognitive effort that directly contributes to learning). Effective multimodal instruction in Aviation English requires balancing these loads—simplifying technical complexity when possible, eliminating redundant or distracting elements, and designing tasks that foster meaningful engagement. For example, integrating video, audio, and text simultaneously can be highly effective, but if poorly sequenced, it may overwhelm learners and hinder vocabulary acquisition.

3.4. Conceptual Framework and Study Limitations

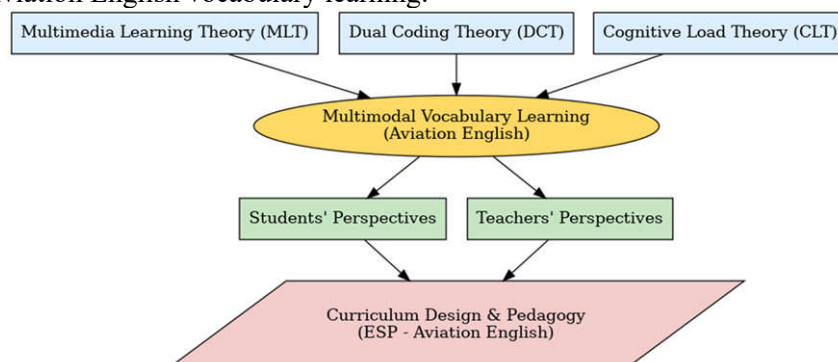
Taken together, these three theories provide a robust framework for designing and evaluating multimodal vocabulary instruction in Aviation English. MLT explains the value of combining visual and verbal inputs, DCT highlights how dual coding reinforces memory and transfer, and CLT provides guidance on managing the cognitive demands of multimodal learning.

For the present study, this integrated framework informs both the design of the survey instruments and the interpretation of findings. Specifically, it guides the analysis of (1) how learners perceive the benefits of multimodal input, (2) whether dual coding enhances their vocabulary retention, and (3) how instructional design can balance effectiveness with cognitive manageability from the learner's perspective.

This study focused solely on students' perspectives and did not capture instructors' views. Future research should include both groups providing a more comprehensive understanding of how multimodal vocabulary learning can be effectively implemented in ESP aviation programs.

3.5. Conceptual Model

Below is the conceptual model illustrating the relationship among the three theories and their application to Aviation English vocabulary learning:



4. Research Methodology

4.1. Research Design

This study employed a mixed-methods comparative design, combining survey data and interviews to investigate student and teacher perspectives on multimodal vocabulary learning in Aviation English. The use of both quantitative and qualitative approaches provided a fuller understanding of preferences, perceived effectiveness, and possible mismatches between learning and teaching practices. The design was informed by Multimedia Learning Theory (MLT), Dual Coding Theory (DCT), and Cognitive Load Theory (CLT), which guided instrument development and data interpretation.

4.2. Participants

Participants were recruited from the Vietnam Aviation Academy in Ho Chi Minh City. The student group comprised 266 undergraduates majoring in Aviation English or related programs, spanning all four years of study. Purposive sampling was used, and surveys were administered online to capture a broad range of learner perspectives.

4.3. Instruments

A structured questionnaire was designed to elicit demographic information, perceptions of multimodal strategies (text, text+image, text+audio, video, augmented reality), and ratings of effectiveness and motivation using Likert-scale items. Cognitive load was assessed with items adapted from Paivio (1986) and Sweller (1998). An open-ended question invited additional comments and suggestions.

5. Results and Discussions

5.1. Results

The student survey comprised three sections: (1) demographic information to outline participants' backgrounds, (2) needs and expectations for domain-specific vocabulary acquisition, and (3) perceptions of Aviation English vocabulary instruction from a multimodal perspective, capturing attitudes toward integrating multiple modes of meaning-making.

5.1.1. The students' background

The survey collected responses from a total of 266 participants. As illustrated in Figure 1, the majority of respondents were female (74.8%), while male participants accounted for 25.2%. This reveals a clear gender imbalance in the sample, with female students forming the dominant group. Such a distribution may influence the interpretation of findings, particularly in areas where gender-related differences in perceptions, attitudes, or learning strategies are relevant. Therefore, the predominance of female respondents should be taken into consideration when generalizing the results to a wider population.

In addition, the largest proportion of respondents is situated at the B1 level, with 119 individuals (44.7%), followed by 94 respondents (35.3%) at the B2 level. At the A2 level, there are 49 respondents (18.4%), whereas only 2 respondents (0.8%) have attained the C1 level. A further 2 respondents (0.8%) are classified as "Unknown," as illustrated in the following figure. The assessment of English proficiency was undertaken to determine learners' current levels within the CEFR framework, thereby providing a foundation for the development of appropriate pedagogical strategies, the adaptation of instructional materials, and the evaluation of program effectiveness.

Among the 266 valid responses, the majority of students (50.4%, $n = 134$) reported their major simply as *English Linguistics*. This group corresponds largely to first- and second-year students, who are enrolled under the Core English Language Studies track before declaring a specialization.

By contrast, upper-year students identified more specific majors: *Aviation English* was reported by 34.6% ($n = 92$), *Business English* by 10.5% ($n = 28$), and *Tourism English* by 1.5% ($n = 4$). In addition, 2.6% ($n = 7$) reported *Aviation Operation Management*, and 0.4% ($n = 1$) reported *Chinese Studies*.

This distribution reflects the program's academic structure: students in Years 1–2 predominantly associate with the general English Linguistics track, whereas students in Years 3–4 report their declared specialization in Aviation English, Business English, or Tourism English.

Table 1. Distribution of reported majors among respondents ($N = 266$)

Major	Frequency	Percentage (%)
English Linguistics	134	50.4
Aviation English	92	34.6
Business English	28	10.5

Tourism English	4	1.5
Aviation Operation Management	7	2.6
Chinese Studies	1	0.4
Total	266	100.0

5.1.2. Students' needs and expectations regarding domain-specific vocabulary acquisition

Regarding prior exposure to aviation English vocabulary, the majority of respondents ($n = 169$, 63.5%) reported that they had never studied or been exposed to such vocabulary. Meanwhile, 32 students (12.0%) indicated that they had studied it before, 63 students (23.7%) were currently studying it, and only a very small proportion ($n = 2$, 0.8%) reported regular use.

When asked about preferred modes of learning Aviation English vocabulary (multiple responses allowed), most students favored official textbooks ($n = 181$, 68.0%), followed by simulations or situation-based videos ($n = 175$, 65.8%) and illustrated images or signs ($n = 167$, 62.8%). More than half also valued flashcards or vocabulary learning applications ($n = 137$, 51.5%). In contrast, very few reported preferences for communication or practical applications ($n = 5$, 1.9%), direct interaction ($n = 1$, 0.4%), or institutional requirements ($n = 1$, 0.4%).

As for the difficulty of aviation English texts, none of the respondents rated them as “very easy.” Only eight students (3.0%) considered the texts “easy,” while nearly half rated them as “moderate” ($n = 122$, 45.9%). A similar proportion reported the texts as “difficult” ($n = 115$, 43.2%), and a smaller group found them “very difficult” ($n = 10$, 3.9%).

The findings reveal notable patterns in students' engagement with aviation English vocabulary. A particularly high proportion of students (63.5%) reported no prior exposure to this lexical domain, reflecting a considerable gap in their specialized vocabulary repertoire. This suggests that their linguistic competence in aviation-related contexts remains limited, especially in terms of recognizing and using domain-specific lexis.

Based on Figure 1, it is evident that learners display varied orientations toward different modalities of acquiring Aviation English vocabulary, with certain modes emerging as particularly salient in shaping their lexical development.

The three most favored methods were official textbooks (181 responses, 68%), simulations or situation-based videos (175 responses, 65.8%), and illustrated images or signs (167 responses, 62.8%). This suggests that while learners value structured materials like textbooks, they also seek visual and practical approaches, with a blended use of these methods offering both theoretical grounding and real-world applicability.

Flashcards and vocabulary learning apps also attracted considerable interest (137 responses, 51.5%), underscoring the role of mobile and flexible tools in supporting vocabulary review and reinforcement, especially for busy learners.

In contrast, direct communication forms received minimal interest—communication/practical applications (5 responses, 1.9%) and direct communication (1 response, 0.4%). Similarly, institutional requirements (1 response, 0.4%) were not viewed as significant drivers of vocabulary learning, suggesting limited awareness or access to communicative practice.

These findings indicate that learners' engagement with Aviation English vocabulary reflects a preference for multimodal input that combines textual, visual, and digital resources. At the same time, the very limited orientation toward communicative forms highlights a linguistic gap in the development of interactional competence, suggesting that while receptive and recognition-based lexical processing is prioritized, productive use in authentic discourse remains underexplored.

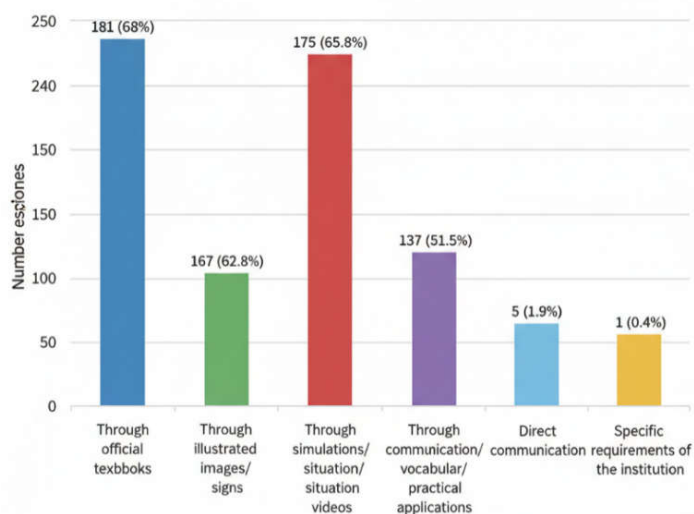


Figure 1: The forms students are interested in learning Aviation Vocabulary

5.1.3. Learning Aviation English Vocabulary through a Multimodal Discourse Approach

The results show that learners benefit most from multimodal input when acquiring specialized vocabulary. Visual images were rated the most effective aid (79.3%), followed by contextual videos and conversational contexts (69.2% each), and written text (67.7%). Audio input (61.3%) supported pronunciation, while gestures (48.1%) and technical symbols (48.9%) played supplementary roles. Overall, the findings highlight that combining multiple modes—especially visual and contextual resources—offers the strongest support for vocabulary retention. These patterns point to a clear linguistic tendency, showing that learners construct meaning more effectively when vocabulary is mediated through multimodal input that combines visual, contextual, and auditory resources.

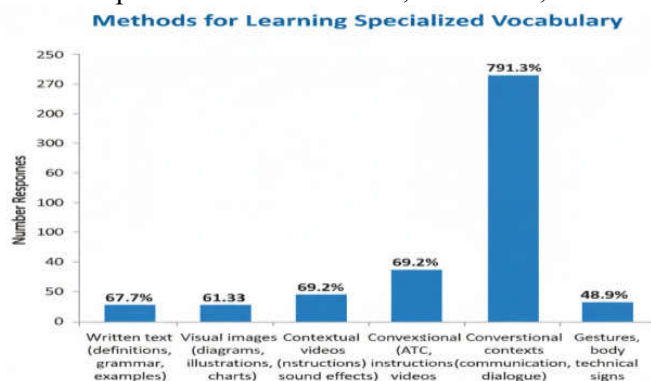


Figure 2. Preferred modalities for learning specialized vocabulary among students' respondents

Responses to “How should aviation vocabulary be presented for easier understanding and retention?” in figure 3 show a strong preference for context-based and multimodal approaches. Real-life or simulated situations were most favored (74.4%), followed by illustrative images (71.8%) and contextual analysis (71.1%). Subtitled videos (68%) and integrated text–audio–image input (65.4%) were also highly valued. Overall, learners view vocabulary acquisition as most effective when grounded in authentic contexts and supported by multimodal resources, rather than through isolated word lists.

The findings indicate that learners process aviation vocabulary most effectively when it is embedded in authentic discourse and supported by multimodal input. The prominence of real-life or simulated scenarios reflects the role of context in facilitating lexical meaning-making, while the strong reliance on visual and video resources underscores the importance of imagery and dynamic input in lexical consolidation. The preference for integrated text–audio–image input further highlights the multimodal nature of vocabulary acquisition, where multiple channels interact to strengthen comprehension and retention.

In figure 4, the results show that most students struggle when learning vocabulary solely from written materials. While 70.7% reported occasional difficulty remembering words and 16.9% found it very difficult to visualize concepts, over 87% in total experienced challenges with text-only input. In contrast, only about 10% preferred this method, and a small minority ($\approx 2\%$) were unsure due to lack of exposure to other approaches.

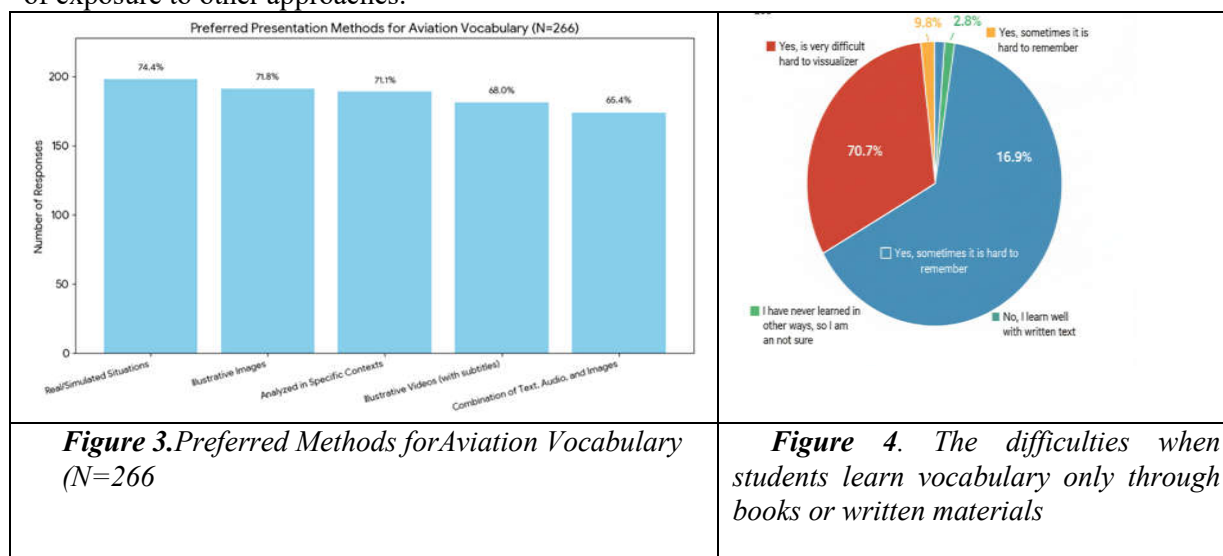


Figure 3. Preferred Methods for Aviation Vocabulary (N=266)

Figure 4. The difficulties when students learn vocabulary only through books or written materials

These findings indicate that while written resources contribute to lexical development, they are insufficient in isolation. Students' reported difficulties with visualization and retention point to the linguistic importance of multimodal input, where the integration of textual, visual, auditory, and contextual elements strengthens the processing and consolidation of specialized vocabulary.

Figure 5 shows that learners prioritize contextual and visual support for vocabulary acquisition, with real-life communication (70.3%) and images (60.9%) rated highest, while auditory input (40.2%), multimodal expression (38%), physical activities (36.5%), and symbols (27.4%) serve as complementary aids.

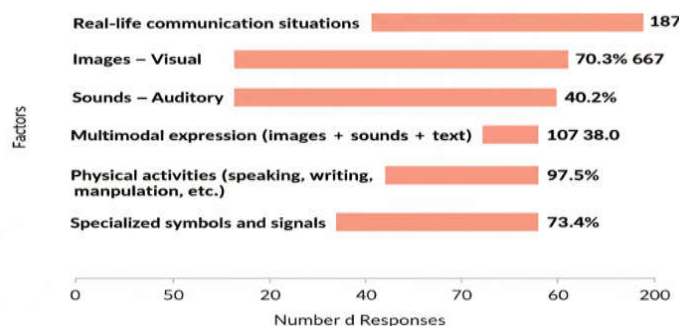


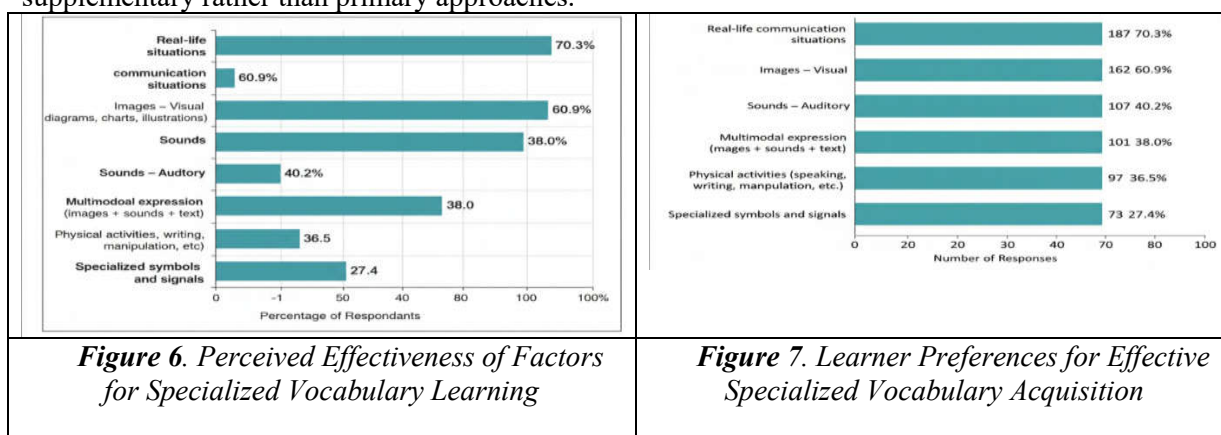
Figure 5. Key Factors Contributing to Effective Specialized Vocabulary Acquisition

The findings indicate that specialized vocabulary acquisition is most effective in context-rich and multimodal environments, where authentic communicative situations provide meaning, visual resources reinforce semantic processing, and the integration of text, sound, and imagery supports retention and usage.

In figure 6, the results for “Which of the following factors do you think help you learn specialized vocabulary most effectively?” indicate that students strongly favor context-rich and visually supported learning. The majority of respondents selected real-life communication situations (70.3%), suggesting that authentic or simulated professional interactions are perceived as the most effective means of acquiring specialized vocabulary. Visual input also played a key role, with 60.9% highlighting the importance of images such as diagrams, charts, or illustrations in supporting comprehension and memory. Auditory factors (40.2%) and multimodal expressions that integrate images, sounds, and text (38%) were also widely recognized, reflecting students’ appreciation for diverse sensory input in reinforcing vocabulary learning. Meanwhile, physical activities like speaking, writing, or manipulation (36.5%) and specialized symbols or signals (27.4%) were selected less frequently, indicating that they are viewed as useful supplementary strategies rather than central approaches.

The findings highlight that specialized vocabulary acquisition is most effective when embedded in authentic communicative contexts and supported by multiple modes of input. Real-life interaction and visual representation emerge as the most salient factors, while auditory and multimodal input further reinforce comprehension and retention. Physical activities and symbols, though less central, function as supplementary resources that enrich the overall linguistic experience.

The survey question “Which of the following factors do you think help you learn specialized vocabulary most effectively?” provides valuable insights into students’ preferred learning supports in figure 7. The most frequently selected option was real-life communication situations (70.3%), underscoring that learners perceive authentic interaction and practice-based contexts as the most effective means of mastering specialized vocabulary. Images were also highly valued (60.9%), indicating that visual reinforcement through diagrams, illustrations, or charts plays a central role in supporting comprehension and recall. Auditory factors (40.2%) and multimodal expression combining images, sounds, and text (38%) were also strongly endorsed, reflecting recognition of the importance of multiple sensory inputs in reinforcing vocabulary acquisition. Physical activities such as speaking, writing, or manipulation (36.5%) and specialized symbols or signals (27.4%) were selected less frequently, suggesting that while these strategies are helpful, students consider them supplementary rather than primary approaches.

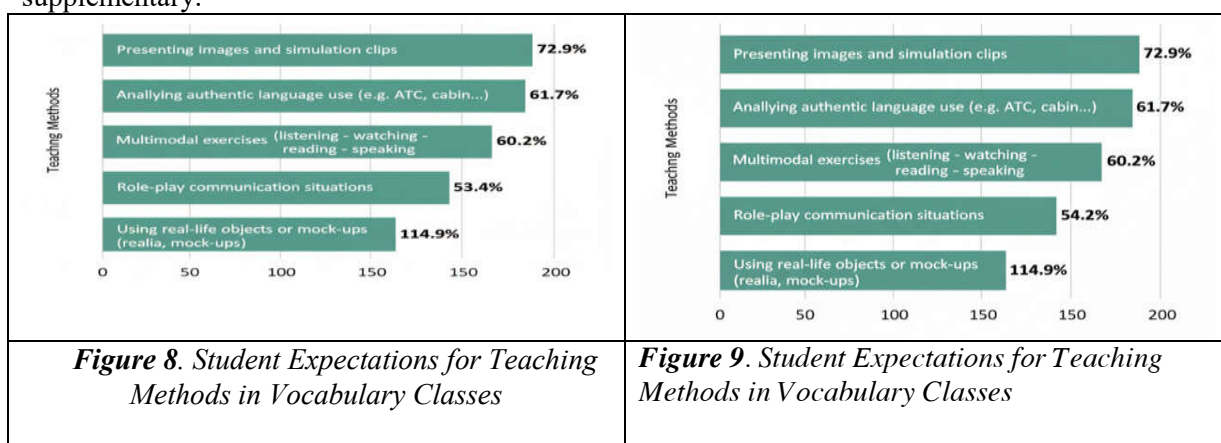


These results highlight that specialized vocabulary is most effectively acquired through context-rich and multimodal input. Real-life communication and visual reinforcement stand out as the strongest supports, while auditory, multimodal, physical, and symbolic resources function as complementary elements that enrich learners' linguistic engagement. The findings indicate that students learn specialized vocabulary most effectively through real-life communication contexts and visual support, while auditory, multimodal, physical, and symbolic inputs serve as complementary aids.

In figure 8, the results of "In vocabulary classes, which methods do you expect the lecturer to use to deliver knowledge?" show that students strongly favor interactive and multimodal teaching approaches. The most preferred method was presenting images and simulation clips (72.9%), highlighting the importance of visual and situational aids in clarifying meaning and improving retention. A majority also valued analyzing authentic language use such as ATC communication or cabin announcements (61.7%) and engaging in multimodal exercises that integrate listening, watching, reading, and speaking (60.2%). These findings emphasize students' preference for exposure to authentic input and integrated learning experiences. Role-play communication situations (53.4%) were also widely endorsed, reflecting learners' recognition of the need for communicative practice that mirrors professional scenarios. Meanwhile, the use of real-life objects or mock-ups (42.9%) received comparatively lower but still significant support, suggesting that tangible materials are considered useful but not as central as visual, contextual, or interactive activities.

The findings indicate that learners view specialized vocabulary acquisition as most effective when supported by authentic input, visual and audiovisual resources, and multimodal engagement. The preference for simulation clips, authentic discourse analysis, and multimodal tasks reflects a tendency toward contextualized, practice-oriented learning rather than rote memorization. Support for role-play highlights the value of experiential use of language in realistic settings, while the presence of realia and mock-ups points to the supplementary role of tangible objects in reinforcing meaning. Overall, the data underscores that vocabulary learning is conceptualized as a multimodal, context-driven process.

Figure 9 shows that students strongly favor context-rich and multimodal methods in vocabulary classes. Presenting images and simulation clips (72.9%) was most preferred, followed by authentic language analysis (61.7%) and multimodal exercises (60.2%). Role-play (53.4%) also reflects the value of experiential learning, while real-life objects or mock-ups (42.9%) were viewed as supplementary.



The findings show that learners conceptualize vocabulary acquisition as most effective when linked to multimodal and context-rich input. The strong preference for simulation clips, authentic discourse, and integrated tasks reflects the value of combining textual, visual, and auditory channels in processing specialized lexis. Support for role-play further illustrates the importance of interactive meaning-making, while the presence of realia and mock-ups indicates the role of tangible reference in reinforcing lexical concreteness. Overall, the results highlight that learners associate effective vocabulary learning with multimodal representation and authentic communicative contexts.

Responses to “Have you ever studied ESP using multimodal methods?” reveal limited exposure. Fewer than a quarter (21.1%) reported prior experience, 33.5% had partial exposure, while nearly half (45.5%) had never engaged with multimodal instruction. These figures show a clear gap between the recognized effectiveness of multimodal techniques and their actual use, suggesting that although such methods are valued in theory, their practical application in ESP contexts remains restricted.

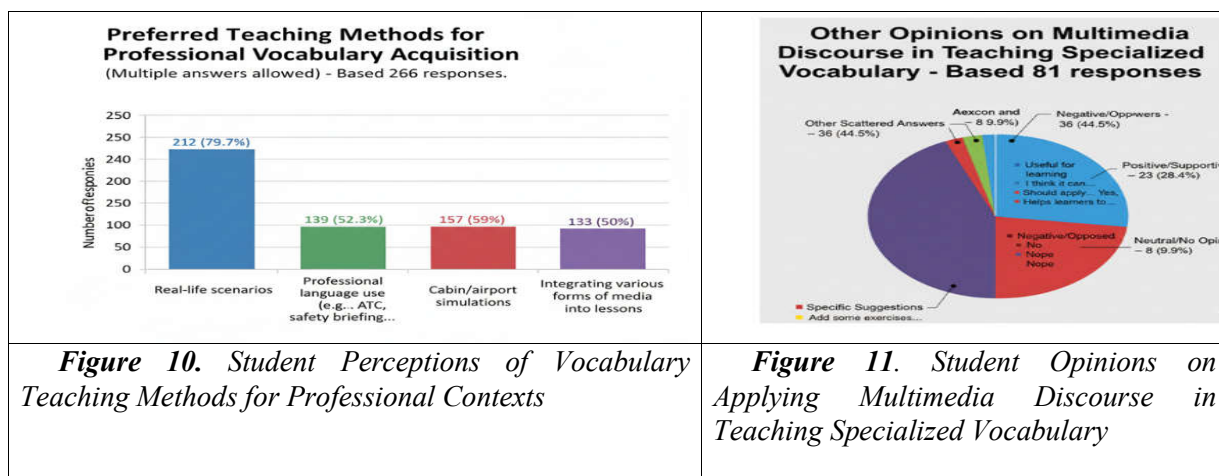
From a linguistic perspective, these findings reveal both a gap and a potential avenue for development. The limited prior exposure suggests that many learners have had little opportunity to engage with multimodal input, which can enhance visualization, contextualization, and retention of specialized vocabulary. At the same time, the data point to the relevance of multimodal resources—such as video performances, simulated dialogues, and visual symbols—as tools that enrich learners’ interaction with language. Their integration may not only bridge current limitations but also support more authentic engagement with professional discourse in ESP contexts.

In figure 10, the results of “Which methods should be used to teach vocabulary for professional contexts?” reveal a clear preference for authentic, practice-oriented approaches. Real-life scenarios were most favored (79.7%), highlighting the importance of situational learning that mirrors professional tasks. Simulations of cabins and airports (59%) further underscore the value of experiential methods in linking classroom learning to workplace realities. Over half endorsed professional discourse such as ATC communication and safety briefings (52.3%), reflecting the necessity of domain-specific input for meaningful vocabulary acquisition. Finally, 50% supported the use of multimodal resources, emphasizing their role in engagement and comprehension.

The findings highlight learners’ preference for vocabulary learning embedded in authentic discourse and professional contexts. Real-life scenarios and simulations emphasize the central role of situated interaction in lexical acquisition, while recognition of professional registers such as ATC communication and safety briefings underscores the importance of domain-specific input for accuracy and meaning. Support for multimodal resources further reflects the value of integrating textual, auditory, and visual channels to enrich comprehension and retention. Overall, students favor approaches grounded in authenticity and practice, combining real-life contexts, simulations, professional language use, and multimedia input.

A survey of 266 students yielded 81 responses (30.5%), a modest rate that limits generalizability but still provides useful insights. Opinions on applying multimedia discourse in specialized vocabulary teaching were mixed. Only small proportions explicitly endorsed it as useful (3.7%), recommended its application (3.7%), or highlighted specific benefits (8.6%). Meanwhile, 11.1% offered cautious approval (e.g., I think it can...), suggesting interest tempered by uncertainty.

In contrast, some students expressed hesitation or resistance: 9.9% gave no opinion, while explicit disagreement appeared in responses such as No (4.9% and 7.4%) and Nope (2.5%), together representing about one-quarter of the total. The largest share (44.5%) provided scattered answers outside defined categories, including practical suggestions like adding exercises (2.5%), which reflect an interest in linking multimedia with applied learning activities.



Overall, the results suggest a mixed yet generally positive view of multimedia discourse, with varying degrees of enthusiasm for its application. Given the low response rate, however, the findings should be interpreted cautiously, as they may not represent the wider student body. Broader participation or complementary qualitative methods would strengthen their reliability and generalizability.

5.2. Discussions

5.2.1. Comparison of Survey Findings with Theoretical Frameworks

The survey reveals students' preference for multimodal, context-rich methods, especially visuals, videos, and simulations. This aligns with Multimedia Learning Theory, as visual (79.3%) and contextual input (74.4%) are seen as key to reinforcing meaning and retention.

The findings align with Dual Coding Theory (Paivio, 1986), showing that combining text, audio, images, and videos enhances memory and recall, with strong support for images (71.8%) and multimodal input (65.4%).

The survey findings reflect Cognitive Load Theory (Sweller, 1998): learning solely through text increases memory strain, while poorly designed multimedia can also overload working memory. This highlights the need for carefully balanced multimodal approaches to optimize learning.

The survey and theories converge on three points:

MLT: verbal-visual integration boosts comprehension.

DCT: multimodal input strengthens retention.

CLT: text-only learning is limited; design must avoid overload.

Overall, vocabulary learning is most effective when multimodal resources are embedded in authentic contexts, combining text, audio, images, and video for stronger engagement and retention.

5.2.2. Comparison with International Research

The findings align with international research on multimodal vocabulary learning:

Li et al. (2022): text + image + audio aid learning but risk overload if poorly designed, echoing student concerns here.

Huang et al. (2022): video-based input improves retention, consistent with students' preference for video.

Godfroid et al. (2017): multimodal support (images, glosses) draws attention to new words, paralleling the emphasis on visuals in aviation terminology.

In Vietnam, limited institutional support for digital tools constrains the use of richer multimodal resources like AR. This gap between theoretical potential and practical feasibility highlights broader challenges of technology-enhanced language learning in developing contexts.

6. Conclusions

This study explored Vietnamese undergraduates' perspectives on multimodal vocabulary learning in Aviation English through the lens of Multimedia Learning Theory (MLT), Dual Coding Theory (DCT), and Cognitive Load Theory (CLT). Drawing on survey responses from 266 students, it revealed a clear preference for multimodal and context-based approaches, emphasizing how learners perceive visual, auditory, and contextual input as central to effective acquisition of specialized vocabulary.

This study extends the application of MLT, DCT, and CLT to the Vietnamese Aviation English context, showing their value in explaining learners' responses to multimodal input. The findings highlight the central role of visual, auditory, and contextual resources in shaping vocabulary acquisition and contribute to ESP research by shifting attention toward learners' perspectives on multimodal vocabulary learning.

In conclusion, this study shows that students strongly favor multimodal and context-based approaches to Aviation English vocabulary learning, with visuals, real-life scenarios, and audiovisual resources seen as most effective. The findings also reveal the limits of text-only input, highlighting the importance of integrated, balanced resources. While this study centers on learners' perspectives, future research should include teachers' views to provide a fuller picture of how multimodal input shapes vocabulary acquisition in Aviation English.

REFERENCES

1. Do, T. T., Nguyen, T. K. T., & Huynh, T. (2024), The evaluation of ESP textbook *Aviation English for ICAO Compliance* from lecturers' perspectives at Vietnam Aviation Academy. *VNU Journal of Science: Education Research*, 40(3), 12–25. <https://doi.org/10.25073/2588-1159/vnuer.4929>.
2. Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H.-J. (2017), Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, 21(3), 563–584. <https://doi.org/10.1017/S1366728917000219>.
3. Huang, Y., Zhang, Z., Yu, J., & Liu, X. (2022), English phrase learning with multimodal input. *Frontiers in Psychology*, 13, Article 828022. <https://doi.org/10.3389/fpsyg.2022.828022>.
4. Li, W., Yu, J., Zhang, Z., & Liu, X. (2022), Dual coding or cognitive load? Exploring the effect of multimodal input on English as a foreign language learners' vocabulary learning. *Frontiers in Psychology*, 13, Article 834706. <https://doi.org/10.3389/fpsyg.2022.834706>.
5. Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511811678>.
6. Nguyen, V. P., & Tran, A. N. (2023), Developing a bilingual multimodal electronic dictionary of airport terms using Book Creator. *Proceedings of AsiaCALL2023*. <https://www.scribd.com/docu/838348095/Booklet>.
7. Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
8. Pham, T. K., & Le, T. H. (2022), Needs analysis of air traffic control students for ESP syllabus design. *Journal of Science, Ho Chi Minh City University of Education*, 20(12), 1–15. <https://test.vjol.info.vn/index.php/sphcm/article/view/90459>.
9. Sweller, J. (1998), Cognitive load theory. *Learning and Instruction*, 12(1), 1–3. [https://doi.org/10.1016/S0959-4752\(01\)00014-7](https://doi.org/10.1016/S0959-4752(01)00014-7).

**Tiếp nhận từ vựng hàng không theo hướng đa phương thức:
Góc nhìn người học dựa trên lý thuyết học tập đa phương tiện (MLT),
lý thuyết mã hóa kép (DCT) và lý thuyết tải nhận thức (CLT)**

Tóm tắt: Nghiên cứu này khảo sát nhận thức của sinh viên về việc học từ vựng đa phương thức trong tiếng Anh hàng không, dựa trên lý thuyết học tập đa phương tiện (MLT), lý thuyết mã hóa kép (DCT) và lý thuyết tải nhận thức (CLT). Khảo sát 266 sinh viên tại Học viện Hàng không Việt Nam cho thấy xu hướng rõ rệt đối với các cách tiếp cận đa phương thức và gắn với ngữ cảnh. Hình ảnh, tình huống thực tế hoặc mô phỏng và phân tích ngữ cảnh được xem là hỗ trợ hiệu quả nhất, trong khi văn bản thuần túy bị đánh giá là hạn chế. Hơn 87% người tham gia cho biết gặp khó khăn trong việc ghi nhớ và hình dung từ vựng nếu chỉ dựa vào tài liệu in. Kết quả nghiên cứu khẳng định giá trị của việc tích hợp ngôn ngữ, hình ảnh và ngữ cảnh, góp phần làm rõ cách tiếp thu, ghi nhớ và vận dụng từ vựng chuyên ngành trong giao tiếp chuyên môn.

Từ khóa: Tiếng Anh hàng không; lý thuyết học tập đa phương tiện; lý thuyết mã hóa kép; đầu vào hình ảnh; phân tích ngữ cảnh.