Indonesian Journal of Linguistics
https://doi.org/10.33005/ijl.v2i2.55

Multimodal Literacy and Visual Grammar in ESP Vocabulary Development: Insights from Air Force Training

Tran Hong Lam [1*]
[1] Faculty of Foreign Languages, Academy of Air and Air Defence Forces, Vietnam

ABSTRACT

Background: This action research study investigates the effectiveness of visual grammar and multimodal resources in teaching and learning specialised English vocabulary within the Air Force context, drawing on Kress and van Leeuwen's visual grammar theory and Unsworth's multimodal literacy framework. **Objective:** The research examines how the compositional, interactive, and representational meanings of visual materials impact vocabulary acquisition and retention among Air Force personnel. **Methods:** The study analyses how visual design elements and multimodal resources support learning specialised Air Force terminology through iterative planning, implementation, observation, and reflection cycles. Python was used to automate the evaluation of visual materials against established criteria, providing an objective, data-driven analysis of their effectiveness. Data triangulation is achieved through three primary collection tools: semi-structured interviews with learners, classroom observation checklists focusing on learner engagement with visual materials, and vocabulary achievement tests measuring retention and comprehension. **Results:** The findings reveal how specific visual grammar elements influence vocabulary learning effectiveness, offering systematic criteria for selecting and designing visual materials in Air Force ESP instruction. **Conclusion:** This research contributes practical guidelines for implementing theoretically grounded visual choices in military ESP contexts while advancing understanding of the application of multimodal literacy in specialised vocabulary instruction.

ARTICLE HISTORY
Submitted 22 Feb 2026
Revised 25 Mar 2026
Accepted 26 Apr 2026
Available online 30 May 2026

KEYWORDS
action research;
air force terminology;
ESP vocabularydevelopment;
multimodal literacy;
visual grammar

INTRODUCTION

In recent years, English for Specific Purposes (ESP) has gained increasing importance in the training of military personnel, especially in fields like aviation, where proficiency in English is crucial for effective communication and safety. Knowing a word's most basic level in learning vocabulary is knowing its form and meaning (Thornbury, 2002). However, traditional methods, such as rote memorisation and translation, have proven inadequate in helping learners retain and use specialised vocabulary effectively (Juhary, 2013). Consequently, new approaches to language instruction are being explored, with multimodal teaching strategies particularly promising. These methods integrate visual, auditory, and kinesthetic modes into the learning process, offering a more engaging and effective alternative to traditional techniques (Unsworth, 2008; Bateman et al., 2017).

*CORRESPONDING EMAIL lamth.1985@gmail.com

Despite the growing use of multimodal teaching in general language-learning contexts, limited research has been conducted on its application in military-specific settings, particularly for teaching ESP vocabulary in aviation and other military fields (Juhary, 2013; Tran and Nguyen, 2024). Visual resources, a key component of multimodal literacy, have been shown to enhance learning by making abstract concepts more accessible and improving vocabulary retention (Kress & van Leeuwen, 2006; Jewitt, 2009). This study aims to investigate the effectiveness of visual strategies in teaching ESP vocabulary to Air Force students, with a particular focus on how visual grammar and multimodal elements can improve both acquisition and retention of specialised terms. The research will also assess students' attitudes toward integrating visual resources into their learning. While traditional teaching methods often prevail in military contexts, incorporating visual elements can encourage greater student engagement and improve students' perceptions of the learning process.

Visual grammar, introduced by Kress and van Leeuwen (2006), provides a systematic approach to understanding meaning through visual modes. Unlike traditional grammar, which focuses on language structure, it examines how images, diagrams, and other visual resources convey meaning. This approach is particularly valuable in ESP contexts like the Air Force, where visual tools significantly support specialised vocabulary learning. Building on this foundation, Jewitt (2009) emphasised multimodal theories, arguing that integrating text, images, and sound creates a richer process of meaning-making. Similarly, Bateman et al. (2017) explored the relationship between visual and verbal modes, underscoring the importance of integrating these modes effectively. Thus, these contributions highlight the critical role of visual grammar in enhancing ESP learning outcomes.

Kress and van Leeuwen (2006) identify three key components of visual grammar: compositional, interactive, and representational meaning, as illustrated in Figure 1. These elements form the backbone for analysing and interpreting visuals in ESP vocabulary teaching. For example, compositional meaning refers to how visual elements are arranged and organised. In ESP contexts such as aviation, the layout of diagrams, like those showing parts of an aircraft engine, helps learners connect vocabulary terms to real-world applications. On the other hand, interactive meaning explores how visuals engage viewers by directing attention through cues such as arrows or gaze direction. These interactive elements help learners focus on essential vocabulary terms in diagrams or charts. Finally, representational meaning concerns how visuals represent real-world objects or concepts.

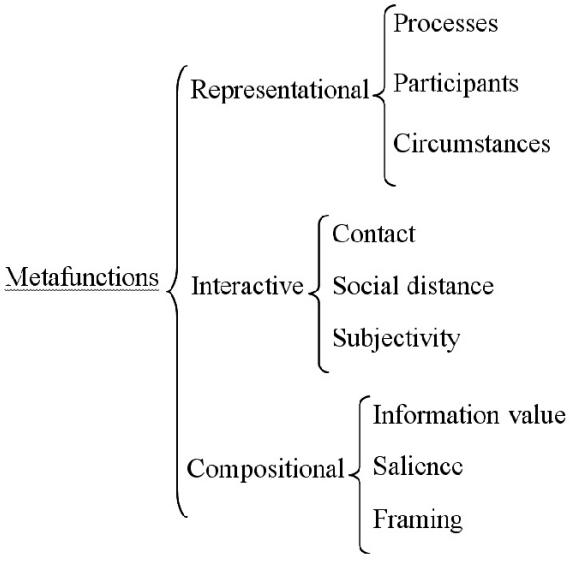

Figure 1. Kress and van Leeuwen's (2006) metafunctional framework

Unsworth's multimodal literacy framework (2001) further complements Kress and van Leeuwen's visual grammar by advocating for integrating multiple modes, text, images, sounds, and gestures into a cohesive learning experience. This integration significantly enhances vocabulary teaching in ESP contexts like aviation, offering learners multiple pathways to connect specialised vocabulary with real-world contexts. For instance, a diagram of an aircraft engine, when paired with textual explanations and interactive simulations, reinforces the vocabulary's visual and conceptual aspects. Unsworth (2008) further extends this framework by including auditory and gestural elements, providing learners with diverse ways to engage with the content. This multimodal approach fosters active engagement by prompting learners to interact with various visual, auditory, and textual modes, thereby enhancing comprehension and supporting more effective vocabulary acquisition.

Choosing appropriate visuals is vital for effective vocabulary instruction in ESP. To that end, the multimodal literacy framework, alongside Kress and van Leeuwen's visual grammar theory (2006) and Unsworth's multimodal framework (2001), offers a structured method for evaluating visuals. These frameworks highlight key criteria for selecting visuals that improve vocabulary acquisition: clarity and organisation, relationships between elements, engagement with the viewer, contextual relevance, accuracy, representation of specialised vocabulary, information value, salience, and framing.

First and foremost, clarity and organisation are essential; visuals must be well-structured with clear labels and layouts, enabling learners to easily connect the information to practical applications (Kress & van Leeuwen, 2006). Similarly, the relationship between visual elements is crucial, as interconnected components help learners link vocabulary to visual cues. Without this connection, poorly connected visuals disrupt the learning process. In addition, visuals should engage viewers through interactive elements, such as arrows or clickable components, that direct attention to key terms and encourage active participation. Furthermore, contextual relevance and accuracy are vital; visuals must accurately represent specialised vocabulary in real-world settings. Misleading visuals can hinder understanding. Moreover, visuals must effectively represent specialised vocabulary, such as connecting terms like "thrust" and "drag" to specific components of an aircraft engine. The information value, such as organising visuals into categories like "given/new" or "ideal/real," helps prioritise key vocabulary elements and supports comprehension. Lastly, salience ensures that key visual elements stand out through techniques such as colour, size, and contrast, while framing maintains the coherence of visual components.

This study applies the visual evaluation framework developed by Tran et al. (2025), which was initially constructed on the basis of Kress and van Leeuwen's (2006) visual grammar theory and Unsworth's (2001) multimodal literacy framework. The framework comprises eight criteria for assessing visuals used in ESP vocabulary teaching: (1) clarity and organisation, (2) relationship between elements, (3) engagement with the viewer, (4) contextual relevance and accuracy, (5) representation of specialised vocabulary, (6) information value, (7) salience, and (8) framing. Each criterion is rated on a 3-point scale, from 1 (low) to 3 (high), resulting in a total score ranging from 8 to 24. Scores from 20–24 indicate high suitability, 12–19 suggest moderate suitability, and below 12 imply that the visual requires revision. This framework was used in the present study to guide the selection and evaluation of visual materials for Air Force ESP instruction. A complete rubric description, including all eight criteria and scoring levels, is available in Table 1 of Tran et al. (2025).

In this study, the visual selection criteria described in Tran et al. (2025, Table 1) were implemented using a Python-based image analysis program. Python's robust libraries: OpenCV, Pillow, and NumPy, were used to automate the evaluation of visuals against the provided criteria, including clarity, salience, and framing. This program, developed and described on *https://github.com/lamth1985/NEUCON-python-based-image-analysis*, was used to analyse a sample set of visuals to evaluate their effectiveness in supporting ESP vocabulary acquisition. Python enables

efficient, objective analysis of image clarity, colour prominence, and the relationships among visual elements (Dixit, 2024). For instance, image sharpness was evaluated using the Laplacian variance method, which quantifies image sharpness (Kumar et al., 2016). Additionally, colour salience was analysed using histogram analysis in the HSV colour space, providing insights into how colour emphasises key vocabulary elements (Hua et al., 2019).

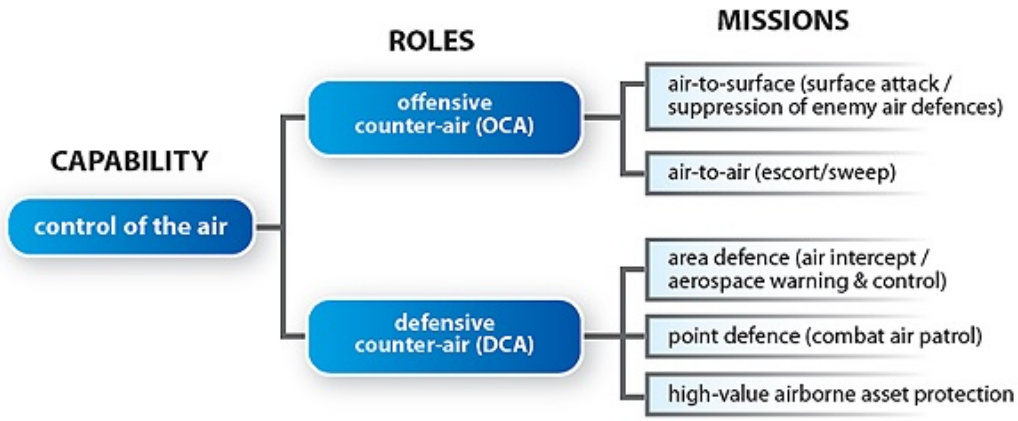

Figure 2. The aerospace control-of-the-air capability (authors' design based on doctrinal sources).

For example, three sample images were analysed using the 3-point scale established in Tran et al. (2025), which was run in the Python program using the code at the GitHub link. This scale evaluates visuals based on their effectiveness in supporting vocabulary acquisition in ESP contexts. Figure 2 scored 17/24, placing it in the "moderately suitable" category. While the diagram was well-structured, some areas could benefit from additional clarity, and the colours were prominent but could have been further enhanced to improve engagement.

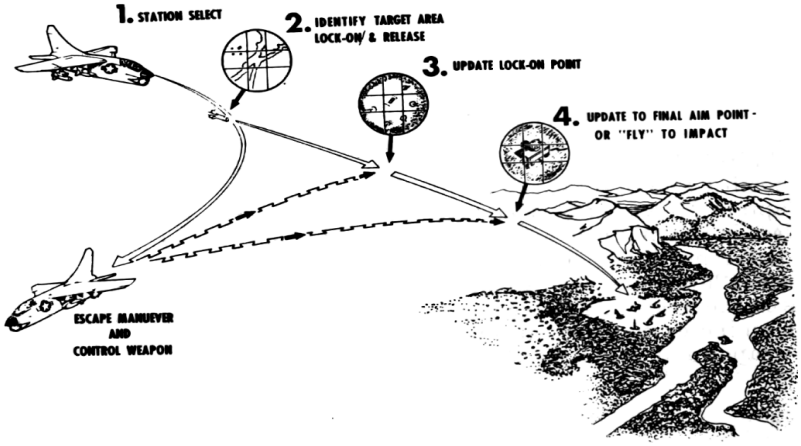

Figure 3. Extended Range Walleye with Data Link (ER-DL) employment standoff against a hostile SAM site. Source: "Fiscal year 1976 and July–September 1976 transition period authorisation for military procurement," testimony of Capt. Zygmont J. Kowalskey, Program Manager, Condor Missile System/Walleye Guided Bomb, Naval Air Systems Command, Part 9, p. 4966. Image via Wikimedia Commons. Public Domain (U.S. Government Work).

Additionally, Figure 3 scored 11/24, indicating it requires revision. The diagram lacked clarity, making it challenging for learners to connect vocabulary to real-world applications. The colour salience was moderate, but not all key elements stood out sufficiently. Figure 4 scored 20/24, making it highly

suitable for vocabulary instruction. The diagram was clear, well-organised, and logically laid out, allowing learners to easily connect vocabulary terms to real-world applications. The colour effectively highlighted key elements, making it engaging and suitable for teaching. These results highlight the effectiveness of Python in evaluating visual elements for clarity and salience, ensuring visuals are optimised for teaching vocabulary in ESP contexts.

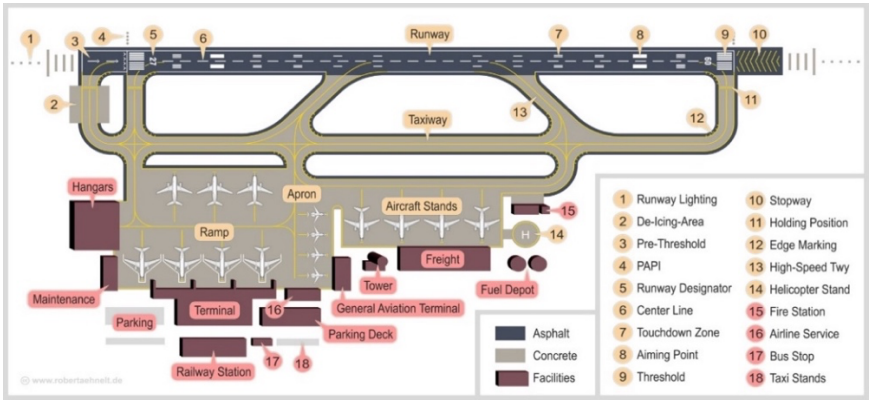

Figure 4. Airport architecture. Source: Aerovertex (https://www.aerovertex.us/), licensed under CC BY 4.0

In conclusion, integrating visual grammar and multimodal literacy provides a powerful approach to ESP vocabulary instruction. By focusing on compositional, interactive, and representational meanings, educators can select visuals that engage and enhance the learning process. These visuals create a richer, more dynamic learning experience, preparing learners for real-world applications combined with text and interactive elements. To ensure objective evaluation, Python was used to automate visual scoring, minimising subjective bias; however, the authors manually curated the selection of images relevant to the lesson content to ensure alignment with the learning objectives.

Recent research on Visual Strategies in ESP teaching highlights the growing interest in incorporating visual elements to enhance vocabulary acquisition, particularly in military and technical contexts. Studies consistently emphasise the importance of visual aids, such as posters, concept maps, and diagrams, in fostering a deeper understanding of specialised terminology. For example, Boștină-Bratu et al. (2018) discuss how posters function as semiotic and linguistic tools that encourage students to decode meanings and engage in critical thinking. This aligns with Kress and van Leeuwen's visual grammar theory, which advocates the use of compositional and interactive visual elements to direct learners' attention and improve vocabulary comprehension.

Moreover, studies by Kulamikhina et al. (2021) and Narimbetov and Azerbaeva (2022) show that integrating Visual Literacy (VL) into ESP courses significantly improves students' vocabulary retention and engagement. These visual strategies support memory and cognitive processes, fostering critical thinking and creativity. The multimodal literacy framework (Unsworth, 2001) further reinforces this by emphasising the integration of multiple modes: text, visuals, and audio to enhance the learning experience and improve the retention of specialised vocabulary.

While these studies demonstrate the effectiveness of visual strategies across general ESP contexts, there remains a significant gap in research on military-specific ESP, particularly in fields such as aviation and air defence. Although studies such as those by Zhang and Liu (2024) and Salazar Pico and Mendoza Bravo (2024) have explored visual learning in technical subjects, they often overlook military training environments, where visual and technical vocabulary are crucial. Thus, our research aims to address this gap by investigating how visual grammar and multimodal literacy can be applied explicitly to ESP military vocabulary training.

METHOD

This section outlines the research design, participants, and data collection methods used in this study. Given that this research is action-oriented and focuses on the effectiveness of visual strategies in teaching ESP vocabulary, the methodology employs a mixed-methods approach, combining qualitative and quantitative data to provide comprehensive insights into the research questions.

This study was conducted during the first semester of the 2024-2025 academic year at the Vietnam Academy of Air and Air Defence Forces, with 31 fourth-year trainees enrolled in the "English for Air Force Combat Training and Navigation" course. This ESP curriculum, designed specifically for military aviation and defence purposes, spans 10 weeks, with 6 class hours per week dedicated to ESP instruction. The course focuses on specialised aviation-related terminology, safety procedures, and technical vocabulary essential for combat training and air navigation within the Air Force context. Additionally, this study was conducted in accordance with institutional and national ethical guidelines. Participation was voluntary, and informed consent was obtained from all participants. All data collected were anonymised and used solely for research purposes.

The participants, assessed at an intermediate (B1) level of English proficiency, have limited prior exposure to visual strategies for learning specialised vocabulary. The course materials include visual aids, but these elements have not been fully explored for their potential to enhance vocabulary retention. This study assesses how visual grammar theory and multimodal literacy frameworks can be integrated into their learning to improve vocabulary acquisition and retention in this military-specific ESP context.

The participants were selected through convenience sampling from the academy's trainee cohort. All participants are male and native Vietnamese speakers. While they possess knowledge of technical vocabulary in their native language, they face challenges in transferring these terms into English due to the complex nature of specialised military terminology. This study seeks to determine whether introducing visual strategies can help bridge this gap, improving participants' ability to recall and apply English military vocabulary effectively in their Air Force training.

For the vocabulary instruction, 30 images were selected to illustrate key terms across eight lessons. These images were sourced from reliable online military educational resources specific to aviation science and military fields from the UK and the U.S., ensuring their relevance to the specialised vocabulary taught. Images were selected based on the criteria outlined in the visual evaluation framework by Tran et al. (2025), ensuring they met the necessary clarity, salience, and relevance standards. The criteria assess the clarity and organisation of visual elements, the relationships among them, and the visual prominence of key vocabulary terms. Once selected, the images were evaluated and scored using the Python program on the GitHub link.

This automated evaluation helped minimise subjective bias, ensuring an objective assessment of the visuals' effectiveness in supporting vocabulary acquisition. The program used Python's OpenCV, Pillow, and NumPy libraries to analyse each image against the set criteria, assigning a score indicating whether the image was highly suitable, moderately suitable, or required revision.

This study adopts an action research design, a cyclical and participatory approach emphasising continuous reflection and improvement in teaching practices. According to McNiff and Whitehead (2002), action research follows a four-step model: Planning, Acting, Observing, and Reflecting. This model is particularly effective in educational settings, as it fosters collaboration between researchers, instructors, and learners, ensuring that teaching strategies evolve based on real-time feedback and outcomes. The study proceeds through three main phases: the Pre-test Phase, Cycle 1, and Cycle 2, each incorporating the four steps to guide the improvement process.

Firstly, in the Pre-test Phase, traditional teaching methods will be employed, without multimodal strategies, to establish a baseline for students' vocabulary knowledge. The pre-test will assess learners' initial understanding and recall of specialised vocabulary. During this phase, we will also plan the multimodal intervention. Specifically, visual grammar theory will inform the selection of visual materials (e.g., images, diagrams, and concept maps) to teach specialised Air Force vocabulary. The careful selection of visuals aims to engage students effectively and help them relate vocabulary terms to real-world contexts.

Secondly, in Cycle 1, multimodal teaching strategies will be introduced, using visual grammar to select and apply visuals such as concept maps, images, and diagrams. These resources will be integrated into four lessons to enhance students' engagement and understanding of specialised Air Force terminology. After completing this phase, a post-test will be administered to evaluate immediate vocabulary acquisition. This phase includes both the acting and observing stages of action research, where we implement the intervention and monitor how students interact with the visual materials. Additionally, semi-structured interviews with instructors and students will gather qualitative feedback, while vocabulary achievement tests will measure the immediate impact of the multimodal strategies.

Lastly, Cycle 2 will continue using the multimodal approach in another set of four lessons, reinforcing and refining the strategies introduced in Cycle 1. This extended exposure will provide a more comprehensive evaluation of how these strategies impact long-term retention and vocabulary usage. A second post-test will assess improvements in long-term retention and additional vocabulary acquisition. This phase will focus primarily on the reflecting stage, where we analyse data from both cycles to inform adjustments to teaching strategies based on student performance and feedback.

Data Triangulation was employed using semi-structured interviews, classroom observation checklists, and vocabulary achievement tests to ensure comprehensive data collection. The checklist summary was administered only after Cycle 2, allowing us to capture the long-term effects of the multimodal intervention. By doing so, we avoid gathering premature data and ensure we assess the overall impact only after the intervention has been fully implemented.

This study uses a mixed-methods approach to assess the impact of visual strategies on vocabulary acquisition in ESP contexts, combining qualitative and quantitative data through semi-structured interviews, classroom observation checklists, and vocabulary achievement tests to gain insights into participants' experiences and progress.

Semi-structured Interviews

Semi-structured interviews were conducted with six learners, the minimum required for phenomenological investigations (Bernard, 2013), selected based on their performance on the pre-test. Two groups were formed: higher-achieving and lower-achieving students, each with two students. The interviews explored participants' perceptions of visual materials and how these strategies influenced their vocabulary acquisition.

The interviews focused on five key questions, which have been included in the Appendix for reference. These questions were designed to gather feedback on learners' experiences with multimodal resources, such as images and diagrams, and on how these materials impacted their learning. Interviews with instructors also explored their experiences with visual grammar theory and its role in improving vocabulary acquisition. All interviews were audio-recorded, transcribed, and analysed for recurring themes such as engagement, learning effectiveness, and preferences for visual materials (Braun & Clarke, 2006).

Classroom Observation Checklists

Classroom observations assessed students' interactions with the visual materials used in the lessons. Researchers utilised observation checklists to record key behaviours, such as students' participation with the visual resources, target vocabulary use, and non-verbal reactions to the visuals. These checklists were designed to objectively measure the extent of student engagement with the multimodal resources and assess whether their interactions contributed to vocabulary acquisition (Cohen, Manion, & Morrison, 2011). The checklist enabled systematic tracking of the effectiveness of visual strategies in enhancing vocabulary retention.

The seven checklist items focused on different aspects of student engagement and vocabulary use. Each item was rated weekly on a scale from 1 to 5, with 1 indicating minimal engagement and 5 indicating maximum engagement. After 8 weeks of intervention, the average ratings were calculated to assess overall engagement and vocabulary acquisition. These ratings, combined with the vocabulary achievement tests and interviews, provided a comprehensive understanding of the impact of multimodal strategies on student learning.

Validity and Reliability of Instruments

To ensure the validity of the qualitative instruments, both the semi-structured interview protocol and the classroom observation checklist were submitted for expert review to two specialists in Applied Linguistics and ESP pedagogy before data collection. Their feedback informed minor revisions to item wording and question sequencing. The interview questions were subsequently piloted with two volunteer trainees outside the main participant pool to verify clarity and to confirm that the questions elicited the intended types of reflective response. Content validity of the vocabulary achievement tests was established by aligning each item with the lesson-specific vocabulary targets and the learning objectives of the "English for Air Force Combat Training and Navigation" curriculum.

Inter-rater reliability for the classroom observation checklist was established by having two independent observers simultaneously code five classroom sessions. Cohen's kappa coefficient ($K = 0.81$) indicated strong inter-rater agreement (Landis & Koch, 1977), providing confidence in the consistency of the observational data. For the qualitative interview data, trustworthiness was addressed through member-checking (returning transcripts to interviewees for verification of accuracy) and peer debriefing (discussing emerging themes with a colleague not involved in the study), consistent with Lincoln and Guba's (1985) criteria for credibility in qualitative inquiry.

The study employed three vocabulary achievement tests consisting of 40 multiple-choice questions, each worth 0.25 points. The test includes recognition tasks (e.g., matching vocabulary with definitions or images) and production tasks (e.g., using vocabulary in context through short writing or speaking exercises). The test was administered within 45 minutes, including time for the answer sheet. A pre-test assessed the learners' initial vocabulary knowledge, while two post-tests were used to measure improvements after the intervention (Read, 2000). Paired t-tests were used to compare the pre-test and post-test results to assess the effectiveness of the multimodal teaching intervention.

This study employs methodological triangulation by integrating three complementary data sources: vocabulary achievement tests (quantitative), classroom observation checklists (structured observation), and semi-structured learner interviews (qualitative). Each instrument targets a distinct but overlapping dimension of the intervention's impact: tests measure learning outcomes; checklists capture in-class behavioural engagement with visual materials; and interviews access learners' perceptions, attitudes, and self-reported learning strategies. Findings are interpreted by seeking convergence across all three data streams, thereby enhancing the credibility of the conclusions and mitigating the limitations inherent in any single data source (Creswell & Plano Clark, 2018).

RESULTS AND DISCUSSION

Vocabulary Achievement Test Results

Descriptive Statistics

The descriptive statistics provide an overview of the participants' performance on the vocabulary achievement tests (pre-test, post-test 1, and post-test 2), presented in Table 1. These tests measured the students' vocabulary acquisition and retention, and the results show a clear trend of improvement.

The pre-test results show an average score of 5.28, suggesting a baseline level of understanding of the specialised vocabulary before the intervention. After the first cycle of multimodal instruction, post-test 1 results show improvement, with the average score rising to 6.09. This increase is accompanied by a higher variance (0.98), suggesting that the multimodal strategies had varying effects across participants. The post-test 2 results show a further increase in the average score to 6.79, indicating continued improvement after the second instruction cycle.

Table 1. Average scores and standard deviations

Groups	Count	Sum	Average	Variance
Pre-test results	31	163.75	5.282258	0.519758
Post-test 1 results	31	189.00	6.096774	0.981989
Post-test 2 results	31	210.75	6.798387	0.993414

One-Way ANOVA Test Statistics

A One-Way ANOVA was conducted to examine whether the differences among pre-test, post-test 1, and post-test 2 scores were statistically significant. The ANOVA summary in Table 2 shows that there is a significant difference in the mean scores:

Table 2. One-Way ANOVA Test Statistics

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	35.69489247	2	17.84745	21.45847	2.4E-08	3.097698
Within Groups	74.85483871	90	0.83172			
Total	110.5497312	92				

The calculated F-statistic of 21.46 is significantly greater than the critical value, indicating a significant difference between the mean scores of the three tests. The P-value of 2.4E-08 is well below the 0.05 threshold, confirming that the differences in the scores are statistically significant. This supports the conclusion that the multimodal intervention had a meaningful effect on vocabulary acquisition.

Paired Two-Sample T-Test for Means

To further assess changes in vocabulary knowledge, paired t-tests were conducted to compare pre-test scores with post-test 1 and post-test 2 scores. The t-tests evaluate whether the mean difference between the scores is statistically significant.

The results of the comparison between the pre-test and post-test 1 are shown in Table 3. The paired t-test between the pre-test and post-test 1 yielded a t-statistic of -7.577 and a P-value of 1.89E-08, significantly lower than the typical threshold of 0.05. This indicates that the increase in vocabulary

knowledge from the pre-test to post-test 1 is statistically significant, suggesting that the first cycle of multimodal teaching strongly impacted vocabulary learning.

Table 3. t-Test: Paired Two Sample for Means (after cycle 1)

	Variable 1 (pre-test)	Variable 2 (post-test 1)
Mean	5.282258065	6.096774194
Variance	0.519758065	0.981989247
Observations	31	31
Pearson Correlation	0.800332707	
Hypothesised Mean Difference	0	
df	30	
t Stat	-7.577368298	
P(T<=t) one-tail	9.46426E-09	
t Critical one-tail	1.697260887	
P(T<=t) two-tail	1.89285E-08	
t Critical two-tail	2.042272456	

Similarly, for the comparison between the pre-test and post-test 2, the results are shown in Table 4. The t-statistic of -11.561 and the P-value of 1.40E-12 further confirm the statistical significance of the improvement in vocabulary knowledge from the pre-test to post-test 2. This reinforces the conclusion that the multimodal strategies contributed to long-term vocabulary retention.

Statistical analysis of vocabulary achievement tests indicates that visual strategies significantly enhance vocabulary acquisition and retention. One-way ANOVA and paired t-tests confirmed a positive, statistically significant impact.

Table 4. t-Test: Paired Two Sample for Means (after cycle 2)

	Variable 1 (pre-test)	*Variable 2 (post-test 2)*
Mean	5.282258065	6.798387097
Variance	0.519758065	0.993413978
Observations	31	31
Pearson Correlation	0.681989954	
Hypothesised Mean Difference	0	
df	30	
t Stat	-11.56184449	
P(T<=t) one-tail	7.0264E-13	
t Critical one-tail	1.697260887	
P(T<=t) two-tail	1.40528E-12	
t Critical two-tail	2.042272456	

Findings from the Interview Analysis

Interviews were conducted with six students, categorised into three performance levels based on their post-test results: two high-performing students (Student 1 and Student 2), two students meeting expectations (Student 3 and Student 4), and two low-performing students (Student 5 and Student 6).

High-performing students were generally positive about the use of visual materials. When asked about their experience with visuals (Question 1), Student 1 shared, "The images and diagrams create

a strong mental connection to the words, making them easier to recall later." In contrast, Student 2 noted, "Seeing a word and its corresponding diagram helps me visualise its meaning, which makes the word more memorable," but also mentioned that some abstract terms still posed challenges.

For students meeting expectations, the feedback was more mixed. Student 3 answered Question 2: "The visuals help with concrete terms like 'airbase' or 'jet,' but I still struggle with abstract words." Meanwhile, Student 4 expressed more optimism, saying, "The visuals are useful, but I need more practice to connect the words to the images fully. It is getting easier, but I need time."

Low-performing students were more cautious in their assessments but were hopeful for improvement with more practice. In response to Question 3, Student 5 admitted, "I find the visuals somewhat helpful, but I often get confused about how they connect to the words." Similarly, Student 6 noted, "The visuals help a little, but I still prefer traditional text-based methods. With more practice, I believe I can understand them better."

These findings show that visual strategies are seen as helpful. However, their effectiveness varies across performance levels, with lower-performing students needing more time and practice to engage with and benefit from the visual materials fully.

Findings from the Observation Checklist

The classroom observation checklist provided valuable insights into how visual materials affected student engagement, vocabulary use, and comprehension, as shown in Figure 5. The findings suggest that visual strategies significantly contributed to students' learning experience and vocabulary acquisition throughout the study.

In terms of engagement, students were highly involved with the visual materials, with an average score of 4.25. Active engagement was most evident during Weeks 2, 5, and 6, when visual resources were consistently used. However, there were slight fluctuations in engagement during Weeks 3 and 7, which may reflect the natural ebb and flow of attention. Despite this, the overall trend indicates that the visual materials effectively maintained student interest and participation throughout the lessons.

Regarding the use of target vocabulary, the checklist recorded a moderate-to-high frequency, with an average score of 3.88. The results showed that as the study progressed, students increasingly incorporated the newly learned vocabulary into their activities, particularly in Weeks 5 and 8, when the multimodal teaching strategies were fully implemented. This suggests that the visual aids supported vocabulary retention and active use during discussions and tasks.

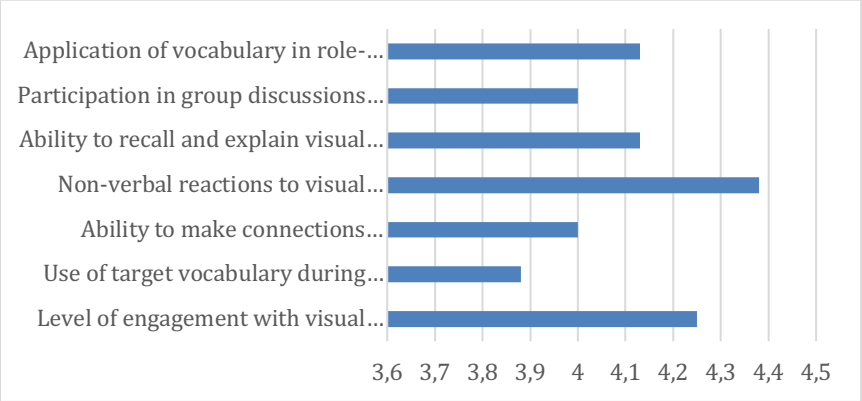

Figure 5. Classroom Observation Checklist Results

Students could also make strong connections between visual elements and vocabulary terms, achieving an average score of 4.00. The ability to relate diagrams and images to specific terms indicates that the visual learning strategies enhanced vocabulary comprehension and retention, enabling students to visualise abstract concepts in real-world contexts.

Overall, the observation findings reinforce the positive impact of visual strategies on vocabulary acquisition and student engagement. The use of visual resources not only facilitated better retention of specialised vocabulary but also encouraged active participation and critical thinking, both crucial in military ESP contexts. These results highlight the effectiveness of multimodal strategies in supporting vocabulary learning, with visual resources proving a valuable tool for training Air Force personnel.

Impacts on Vocabulary Acquisition and Retention

The results from the vocabulary achievement tests confirm that visual strategies were highly effective in improving both vocabulary acquisition and retention. The average score increased from 5.28 in the pre-test to 6.80 in post-test 2, suggesting that visual materials, such as pictures and diagrams, helped students better understand and retain technical vocabulary. This aligns with Kress and van Leeuwen's (2006) Visual Grammar Theory, which posits that the interaction between visuals and text enhances comprehension by linking abstract concepts to concrete representations.

Although the intervention had a positive overall effect, lower-performing students needed additional practice to integrate the visuals into their learning fully. This finding is consistent with Mayer and Moreno's (2003) research, which emphasises that the effectiveness of multimedia learning depends on learners' prior knowledge. High-performing students benefited most from the visual strategy, with improved recall and application of vocabulary in real-life contexts, as demonstrated in group discussions and role-play tasks. Their improved post-test performance reflects the effectiveness of visual materials in supporting vocabulary retention.

Attitudes Toward Visual Learning

Regarding attitudes toward visual strategies, high-performing students responded very positively. In interviews, these students expressed that visuals significantly aided their understanding and retention of complex technical vocabulary. Student 1 remarked, "The images and diagrams create a strong mental connection to the words, making them easier to recall later." This feedback aligns with Boștină-Bratu et al. (2018), who found that visual aids foster engagement and improve comprehension, mainly when dealing with specialised vocabulary.

In contrast, students meeting expectations and those lower-performing had mixed reactions. While some students appreciated the value of the visual aids, others struggled to connect the visuals to vocabulary terms. Student 3 stated, "The visuals help with words like 'jet' or 'airbase,' but I still struggle with more abstract terms." This suggests that while visual strategies are beneficial, their effectiveness varies depending on students' familiarity with the concepts. Lower-performing students reported greater challenges integrating visuals into their vocabulary learning. Student 5 admitted, "I find the pictures and diagrams somewhat helpful, but sometimes I get confused about how the visuals relate to the words." These students indicated the need for more practice and exposure to fully benefit from multimodal learning strategies.

These findings support Duong's (2022) suggestion that visual aids are especially effective when students already have some basic understanding of the content. For lower-performing students, additional time, support, and possibly more explicit guidance on connecting visuals with vocabulary are necessary for effective learning.

In short, this study provides strong evidence that visual strategies significantly enhance vocabulary acquisition and retention among Air Force trainees. Visual materials such as pictures and diagrams proved particularly effective in helping students understand and retain specialised terminology, especially for high-performing students. However, the findings also highlight that lower-performing students struggle to connect visuals to vocabulary, suggesting a need for additional practice. These results contribute to the growing body of literature on the effectiveness of multimodal strategies in ESP teaching, particularly within military contexts. Future research should explore optimising visual strategies for different proficiency levels to ensure all learners benefit from multimodal learning.

Several limitations of this study warrant acknowledgement. First, the sample was small (n = 31) and demographically homogeneous: all participants were male, native Vietnamese speakers at B1 proficiency, and enrolled at a single military institution, which limits the transferability of the findings to other ESP or military training contexts. Second, the action research design did not include a control group receiving conventional instruction. While the pre-test–post-test comparisons (ANOVA, paired t-tests) provide internal evidence of improvement across the intervention cycles, the absence of a control condition means that alternative explanations for score gains, such as maturation, repeated test exposure, or general instructional quality, cannot be fully ruled out. These design characteristics are inherent to single-group action research. Still, they should be addressed in future studies through quasi-experimental designs with matched comparison groups and by involving more diverse participant cohorts across multiple institutions.

CONCLUSION

This study investigated the effectiveness of visual strategies, particularly visual grammar and multimodal resources, in enhancing specialised English vocabulary acquisition among Air Force trainees. Based on classroom observations and vocabulary achievement tests, the findings indicate that integrating visual materials significantly improved vocabulary comprehension and retention. Students engaged more with the visual elements, actively using target vocabulary in discussions and role-play tasks. Statistical analyses confirmed the positive impact, showing significant improvements in test scores and vocabulary acquisition.

Despite these benefits, some students struggled to connect visuals to vocabulary, particularly those with lower proficiency. This suggests that while visual strategies are highly effective, their impact may vary depending on individual learning styles and proficiency levels. Overall, the study supports integrating visual grammar and multimodal resources into ESP instruction, particularly to enhance vocabulary learning and its practical application.

To optimise visual strategies in ESP teaching, educators should ensure that visual materials are explicit, contextually relevant, and consistent. Principles of visual grammar, such as compositional and interactive meaning, should guide the selection and design of visuals to enhance conceptual understanding. Various visual formats, including diagrams, infographics, and interactive elements, should cater to diverse learning styles and reinforce vocabulary retention.

Future research could examine long-term vocabulary retention beyond the immediate post-test window and compare the effectiveness of different visual modalities (e.g., static diagrams vs animated visuals). To address the generalizability constraint of the present study, future investigations should (a) replicate the intervention in mixed-gender or mixed-proficiency military cohorts, (b) adopt quasi-experimental designs with matched control groups receiving conventional instruction to isolate the specific contribution of visual grammar-guided materials, and (c) compare outcomes across different military branches or civilian ESP contexts such as maritime, medical, or aviation maintenance training. Employing stratified purposive sampling would also help ensure

greater participant diversity. Finally, teacher training programmes should incorporate systematic guidance on selecting and designing grounded visual materials theoretically, equipping instructors with practical tools aligned with visual grammar frameworks.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to the Vietnam Academy of Air and Air Defence Forces for their support and for the resources they provided to facilitate this study. We are also grateful to the students who participated and contributed valuable feedback. Our heartfelt thanks go to the Faculty of Aeronautical Engineering and the Faculty of Languages for their intellectual guidance and academic assistance throughout the research process. Finally, we extend our deepest appreciation to our families for their unwavering support and encouragement.

CONFLICT OF INTEREST

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

FUNDING INFORMATION

This research received no external funding and was fully self-financed by the author.

REFERENCES

Bateman, J., Wildfeuer, J., & Hiippala, T. (2017). *Multimodality: Foundations, research, and analysis – A problem-oriented introduction*. Berlin, Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110479898

Bernard, H. R. (2013). *Social research methods: Qualitative and quantitative approaches* (2nd ed.). Los Angeles, CA: SAGE Publications.

Boștină-Bratu, S., Negoescu, A. G., & Palea, L. (2018). Visuality and communication in military English classes. *International Conference Knowledge-Based Organization, 24*(2), 269–275. https://doi.org/10.1515/kbo-2018-0101

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). London: Routledge. https://doi.org/10.4324/9780203720967

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research (3rd ed.)*. Thousand Oaks, CA: SAGE.

Dixit, A. (2024). Basics of image analysis and manipulation using Python. In I. A. Ansari & V. Bajaj (Eds.), *Image processing with Python: A practical approach* (pp. 1-1 to 1-27). IOP Publishing. https://doi.org/10.1088/978-0-7503-5924-5ch1

Duong, T. M. (2022). Insights into ESP vocabulary learning strategies used by Vietnamese tertiary students. *Journal of Language and Education, 8*(1), 38–49. https://doi.org/10.17323/jle.2022.10924

Hua, J.-Z., Liu, G.-H., & Song, S.-X. (2019). Content-based image retrieval using color volume histograms. *International Journal of Pattern Recognition and Artificial Intelligence, 33*(1), 1940010. https://doi.org/10.1142/S021800141940010X

Jewitt, C. (2009). *The Routledge handbook of multimodal analysis*. London: Routledge.

Juhary, J. (2013). Using military vocabulary and concepts for successful language learning at the Defence University. *Literacy Information and Computer Education Journal, 4*(3), 1178–1185. https://doi.org/10.20533/licej.2040.2589.2013.0156

Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). London: Routledge.

Kulamikhina, I., Kamysheva, E., & Samylova, O. (2021). Integrating visual literacy teaching into the English for Specific Purposes course. In O. Kolmakova, O. Boginskaya, & S. Grichin (Eds.), *Language and technology in the interdisciplinary paradigm* (Vol. 118, pp. 689–702). European Proceedings of Social and Behavioural Sciences. https://doi.org/10.15405/epsbs.2021.12.84

Kumar, P., Raj, G., & Choudhury, T. (2016). Blur image detection using Laplacian operator and Open-CV. *Proceedings of the 2016 International Conference on System Modeling & Advancement in Research Trends (SMART)*, 63-67. https://doi.org/10.1109/SYSMART.2016.7894491

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: SAGE.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52. https://doi.org/10.1207/s15326985ep3801_6

McNiff, J., & Whitehead, J. (2000). *Action research in organisations* (2nd ed.). London: Routledge.

Narimbetov, M., & Azerbaeva, E. (2022). Importance of implementing visual aids in teaching ESP students. *Eurasian Journal of Learning and Academic Teaching, 5*, 29–32. Retrieved from https://geniusjournals.org/index.php/ejlat/article/view/624

Read, J. (2000). *Assessing vocabulary* (Cambridge Language Assessment). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511732942

Salazar Pico, I. M., & Mendoza Bravo, K. L. (2024). Visual-spatial learning to enhance teaching vocabulary. *Revista Científica Sinapsis, 24*(1). https://doi.org/10.37117/s.v24i1.1025

Thornbury, S. (2002). *How to teach vocabulary*. Harlow: Longman.

Tran, H. L., & Nguyen, K. L. (2024). Semantic evolution in U.S. Air Force terminology: A corpus-based analysis. In *Proceedings of the International Conference on Multidisciplinary Research ICMR12, 2024* (Vol. 2, pp. 116–128). Agricultural Institute Publishing House.

Tran, H. L., Nguyen, K. L., To, T. L. H., & Vu, X. T. (2025). Multimodal strategies for image selection in Air Force terminology teaching: An interdisciplinary study using visual grammar and Python-based image analysis. In *Proceedings of the 21st International Conference on Socio-Economic and Environmental Issues in Development, 2025* (pp. 3775–3787). Finance Publishing House.

Unsworth, L. (2001). *Teaching multiliteracies across the curriculum: Changing contexts of text and image in classroom practice*. Buckingham: Open University Press.

Unsworth, L. (2008). Multiliteracies, e-literature, and English teaching. *Language and Education, 22*(1), 62–75. https://doi.org/10.2167/le726.0

Zhang, J., & Liu, Z. (2024). A study of vocabulary teaching strategies in ESP instruction and their effects on students' English vocabulary learning using the Subsumption Sorting Algorithm. *Applied Mathematics and Nonlinear Sciences, 9*(1), 1–18. https://doi.org/10.2478/amns-2024-1587