English Language & Literature Teaching, *Vol. 32,* No. 2 Summer 2026
https://doi.org/10.35828/etak.2026.32.2.51

A Study of Video-Based ESP Vocabulary Learning and Retention in a Chinese Vocational University

Xingxue Liu*
(Shandong Vocational University of Foreign Affairs)
Sunmee Chang**
(Hoseo University)

Liu, Xingxue, & Chang, Sunmee (2026). A study of video-based ESP vocabulary learning and retention in a Chinese vocational university. *English Language & Literature Teaching, 32*(2), 51-71.

This study examines the effectiveness of video-based instruction for English for Specific Purposes (ESP) vocabulary learning and retention among 66 Chinese vocational university students (Group A, n = 41; Group B, n = 25). Using a mixed-methods quasi-experimental design over a ten-week period, the research rigorously compared video-based learning and traditional classroom instruction across two curriculum-aligned stages. Quantitative data included pre-tests, immediate post-tests, and delayed post-tests; qualitative data came from student questionnaires and semi-structured interviews. Findings show that video-based instruction significantly supported vocabulary acquisition, particularly for foundational technical terms, with Group A scoring 12.10 (out of 15) on the immediate post-test following video-based instruction compared to 8.86 following traditional instruction. Retention patterns over a two-week interval indicated more stable form-meaning links for concrete vocabulary, though complex terminology remained challenging, with Group B showing a decline from 6.10 to 3.30 in delayed post-tests. Learners reported enhanced clarity, increased motivation, and greater learner autonomy, while also noting limitations in classroom interaction and potential cognitive overload for abstract content. Ultimately, the study provides strong evidence that curriculum-aligned video-based instruction can effectively support ESP vocabulary learning when combined with principled instructional design, though findings are interpreted as condition-based associations due to intact-class constraints.

[ESP vocabulary/video-based instruction/multimedia learning/cognitive load/ vocational education]

---

* First author ** Corresponding author

I. INTRODUCTION

English for Specific Purposes (ESP) is a core component of China's higher vocational education, equipping students in engineering and electronics with specialised terminology for manuals, production protocols, and workplace communication. Despite prior general English study, many vocational learners struggle with ESP vocabulary, particularly long-term retention, as technical terms are tied to abstract concepts and industrial processes, placing heavy cognitive demands on learners who must process linguistic meaning and domain knowledge simultaneously (Coxhead, 2025). In practice, ESP vocabulary teaching relies heavily on teacher-centred explanation and memorisation, introducing large quantities of new terms quickly with limited repetition or contextualised use—practices that often result in superficial lexical knowledge, further constrained by time pressures within vocational curricula.

Video-based instruction, which delivers combined visual and verbal input, offers potential solutions. Videos can make abstract processes more concrete and allow students to pause, replay, and review content at their own pace—features particularly valuable for ESP vocabulary learning, where understanding terminology depends on grasping underlying technical concepts. However, the use of instructional videos for ESP vocabulary learning in Chinese vocational universities remains under-researched. Most prior studies focus on general English or academic contexts, leaving vocational learners under-represented. Few studies examine both acquisition and retention alongside learners' perceptions in authentic classrooms, resulting in limited evidence on whether integrating instructional videos improves vocabulary outcomes.

Drawing on multimedia learning theory (Mayer, 2002) and cognitive load perspectives, this study investigates how curriculum-aligned video-based instruction supports ESP vocabulary acquisition, retention, and learner engagement among vocational university students. It addresses three research questions: (1) What patterns of ESP vocabulary acquisition are observed under video-based instruction? (2) To what extent is ESP vocabulary retained following video-based instruction? (3) How do vocational students engage with and perceive video-based instruction in ESP courses?

II. LITERATURE REREVIEW

1. Challenges in ESP Vocabulary Learning in Vocational Contexts

ESP vocabulary comprises specialised lexis tightly bound to professional contexts (Webb & Nation, 2017), including single technical terms (modulation), multi-word units (analog-

to-digital converter), and semi-technical items. Unlike general vocabulary, ESP items display low frequency, high specificity, and limited semantic transparency, requiring learners to process both linguistic form and disciplinary concepts simultaneously. This places substantially higher cognitive demands on vocational students than general English. Technical items such as photolithography or analog-to-digital conversion encode complex scientific processes and abstract relationships often unfamiliar to learners, requiring simultaneous attention to form, meaning, and disciplinary concepts, thereby increasing intrinsic cognitive load. For vocational learners with limited prior exposure to academic English or technical theory, this burden can easily exceed working-memory capacity (Baddeley, 2010), resulting in superficial learning or rapid forgetting. The Noticing Hypothesis (Schmidt, 1990) and Depth of Processing Theory (Lockhart & Craik, 1990) further explain why traditional instruction often fails: technical terms introduced quickly through dense explanations reduce conscious noticing and promote shallow memorisation rather than meaningful elaboration. Task-based research, including the Involvement Load Hypothesis (Hulstijn & Laufer, 2001), indicates that higher-involvement tasks lead to stronger learning, yet such approaches are rarely implemented in vocational settings.

These cognitive challenges are compounded by instructional and contextual realities. ESP courses often meet infrequently, yet compressed curricula require students to learn large quantities of new vocabulary within limited time. From a Cognitive Load Theory perspective (Sweller, 1988), teacher-centred practices—lengthy verbal explanations, dense textbook descriptions, and minimally segmented input—create substantial extraneous load that competes for limited working-memory resources. Learner-related factors further complicate matters: many vocational students have uneven English proficiency, lack effective strategies for abstract terminology, and may perceive ESP as secondary to core professional subjects, undermining motivation and sustained engagement. Retention poses an additional challenge, as newly learned vocabulary is highly vulnerable to decay without review or reuse (Radvansky et al., 2022). ESP terminology, being low-frequency and rarely recycled systematically, is particularly susceptible to forgetting, yet traditional courses often introduce terms once and move on, with few opportunities for spaced repetition. This study operationalises vocabulary knowledge primarily as form–meaning recognition and productive recall assessed through immediate and delayed tests, focusing on accuracy and durability of form–meaning mapping—a focus that also represents a limitation, as receptive–productive distinctions and deeper aspects (e.g., collocational competence) remain underexamined in vocational contexts (Schmitt & Schmitt, 2014).

2. Video-Based Instruction and Multimedia Learning

Video-based instruction, defined as the use of digital instructional videos deliberately

designed to support curricular learning (Fiorella & Mayer, 2018; West & Graham, 2005), combines spoken language with visual representations, enabling learners to process information through coordinated verbal and visual channels. This aligns with Mayer's Cognitive Theory of Multimedia Learning, which views instructional video as multimodal input intended to enhance learning through integration of words and pictures (Mayer, 2002, 2005). A distinctive affordance is temporal control and replayability: learners can pause, rewind, or rewatch explanations as needed—an option not typically available in real-time classroom interaction. These features are particularly consequential in vocational ESP contexts, where technical vocabulary refers to processes, mechanisms, or symbolic systems difficult to convey through text or static images alone.

The potential effectiveness of video-based instruction can be understood through complementary theoretical perspectives. Dual Coding Theory (Paivio & Clark, 1991) proposes that information is processed in parallel through verbal and visual systems, and learning is strengthened when material is encoded across both channels—particularly relevant in ESP where terms such as *photolithography* benefit from visual representation of underlying processes. Mayer's theory specifies design principles—segmentation, signalling, coherence, and modality—that optimise learning given limited processing capacity. Cognitive Load Theory (Sweller, 1988, 2011) explains how video-based instruction may be advantageous. To clarify the connection between Cognitive Load Theory and the present intervention, video-based instruction addresses each dimension as follows. First, extraneous load is reduced through segmented presentation and concise narration, eliminating lengthy, densely packed verbal explanations typical of traditional instruction. Second, intrinsic load is managed by visualizing abstract technical processes through animations and diagrams, making complex concepts more accessible and reducing the burden of mental visualization. Third, germane load is promoted through learner-controlled pacing and replayability, which allow students to allocate additional cognitive resources to schema construction and deeper integration of form-meaning mappings. This is supported by research showing students engage more effectively with concise, well-structured videos than with long or visually complex recordings (Guo et al., 2014). Pedagogical affordances relevant to vocational ESP include enhanced conceptual clarity through visualisation of technical processes, differentiated self-paced learning enabling students to regulate exposure and revisit challenging sections, opportunities for spaced review and recycling of key terminology, and positive influences on motivation and engagement through visually rich, applied representations of professional practice.

However, the evidence base for video-based vocabulary instruction remains limited. Many existing studies involve small convenience samples, short instructional periods, or minimal exposure to video materials, constraining generalisability. Delayed post-tests are frequently omitted, making it difficult to determine whether learning gains persist beyond the

immediate instructional window. Vocabulary assessments often prioritise recognition-based measures rather than productive or conceptually rich knowledge, and vocational learners remain under-represented compared with general EFL or university students. Moreover, instructional videos used in prior research are sometimes poorly aligned with curricular content, raising questions about ecological validity. Parallel research on microlearning reports positive effects on vocabulary learning efficiency and learner attention (Ghafar et al., 2023), but overly fragmented presentation may hinder integration of complex content (Mayer, 2005). Across both strands, three outcome domains emerge: acquisition (generally positive), retention (mixed, dependent on spacing and repetition), and engagement (robustly positive, though not universal). Few studies focus specifically on ESP vocabulary in vocational university contexts or investigate both acquisition and delayed retention alongside learner-experience data in authentic classrooms. These gaps directly motivate the present study, which examines curriculum-aligned video-based instruction for ESP vocabulary learning in an authentic vocational classroom setting.

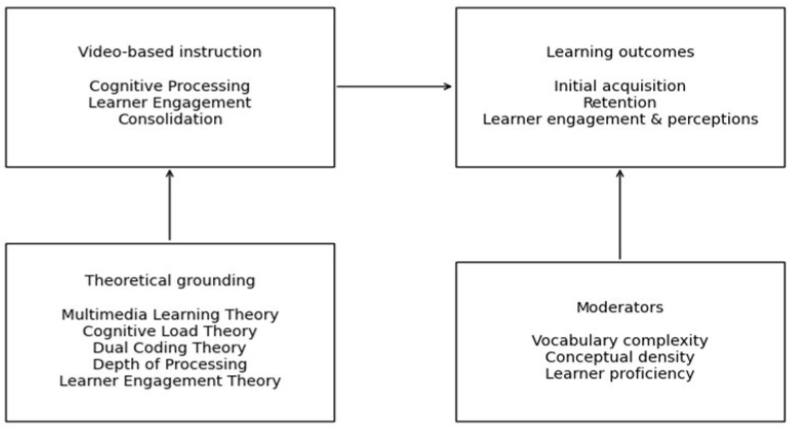
3. Conceptual Framework of the Study

The conceptual framework (Figure 1) integrates theories of vocabulary acquisition, multimedia learning, cognitive load, and learner engagement. Grounded in Mayer's Cognitive Theory of Multimedia Learning and Cognitive Load Theory, complemented by Dual Coding Theory, Depth of Processing Theory, the Involvement Load Hypothesis, and research on learner engagement (Fredricks et al., 2004), the framework conceptualises video-based instruction as mediating the relationship between instructional input and learning outcomes. Three central assumptions guide the framework. First, ESP vocabulary learning is cognitively demanding because learners must process unfamiliar linguistic forms while simultaneously constructing understanding of underlying technical concepts; learning therefore depends not only on exposure but on how information is represented and processed (Nation, 2017). Second, instructional design plays a crucial role: when content is presented through coordinated visual and verbal channels, structured into manageable units, and stripped of unnecessary complexity, learners are more likely to engage in deeper, meaning-focused processing than when instruction relies primarily on dense verbal explanation. Third, learner engagement mediates learning outcomes: the degree to which learners perceive instruction as clear, relevant, and manageable, and the extent to which they can control pacing and review, influences how actively they engage with learning tasks and, consequently, how well they acquire and retain new vocabulary.

Within this framework, video-based instruction—operationalised as curriculum-aligned instructional videos with narrated explanations, visual representations, and segmented content—shapes vocabulary learning through three interrelated processes. First, it supports

cognitive processing by coordinating auditory and visual input, externalising complex technical concepts through diagrams or animations, and regulating cognitive load through segmentation and signalling. Second, the clarity, relevance, and self-paced nature of instructional videos enhance learner engagement and autonomy, particularly in vocational classrooms characterised by mixed proficiency levels. Third, because videos can be revisited, they provide opportunities for repetition and consolidation, enabling distributed review and supporting the stabilisation of form–meaning mappings in memory. These processes link instructional input to three principal outcome domains: initial ESP vocabulary acquisition (learners' understanding and recognition or production of new technical terms immediately after instruction), ESP vocabulary retention (the durability of that knowledge over time), and learner engagement and perceptions (cognitive, behavioural, and affective responses to video-based instruction). Vocabulary complexity, conceptual density, and learner proficiency operate as moderating influences. A distinctive feature of the framework is its curriculum-aligned orientation: vocabulary instruction is embedded within an existing ESP syllabus, allowing learning to be examined within coherent instructional sequences while preserving ecological validity and alignment with normal course progression.

FIGURE 1
Conceptual Framework for the Study



III. RESEARCH DESIGN AND METHODOLOGY

1. Research Design, Context and Participants

This study adopted a mixed-methods quasi-experimental design in an authentic vocational

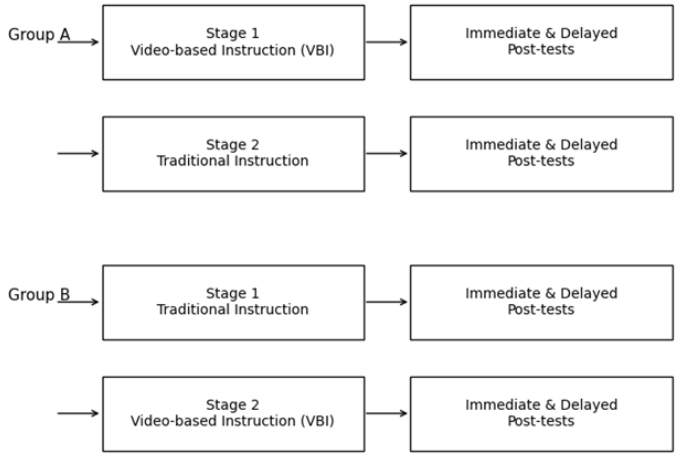
ESP classroom at a university in eastern China, where ESP courses are compulsory for technical majors and traditionally follow a teacher-centered model focused on vocabulary explanation, bilingual clarification, and recognition-based exercises. Two intact classes participated over ten weeks following the same curriculum sequence. As summarised in Table 1, Group A comprised 41 second-year IoT students; Group B consisted of 25 first-year Applied Electronics students. Both groups had no prior experience with video-based ESP instruction. Because random assignment was not feasible and the groups differed in baseline proficiency (reflected in pre-test *M*), the study adopts a parallel-group orientation emphasising within-group learning trajectories rather than between-group comparison. Instruction was delivered across two curriculum-aligned stages (Weeks 2–6 and 6–10), with each group experiencing both video-based and traditional instruction in reversed order. However, the vocabulary sets in each stage were drawn from different curriculum units and differed in content and conceptual demands; no assumption of equivalence is made, and the sequence is not treated as a crossover design. Acquisition and retention are therefore interpreted as within-group, condition-based observations. Participants were recruited through convenience sampling. In Group B, five students were absent on the pre-test administration day due to scheduling conflicts with other course examinations, resulting in a valid pre-test sample of 20 for Group B (original $n = 25$). All 25 students participated in subsequent assessments when present. Exclusion was applied only for missing assessments or withdrawal. Within this framework, quantitative vocabulary data track learning across stages, while questionnaire and interview data illuminate learners' engagement and perceptions. The researcher served as the instructor for both conditions, which may introduce teacher effects; to mitigate this, both conditions followed identical structured lesson plans.

The overall design is summarised in Figure 2, which illustrates the ten-week instructional sequence including pre-test, two instructional stages with reversed conditions across groups, immediate post-tests after each stage, two-week retention intervals, delayed post-tests, and final questionnaire and interviews. As shown in Figure 2, both groups followed the same assessment timeline while receiving different instructional modes at each stage, allowing for within-group, condition-based observations

TABLE 1
Participant Demographics

Group	*n*	Male/Female	Age Range	Year
Group A	41	21/20	17–20	2nd
Group B	25	11/14	17–20	1st
Total	66	32/34	17–20	—

FIGURE 2
Overview of the Study Design



2. Instructional Materials, Procedures and Instruments

The video-based condition consisted of four short, curriculum-aligned videos (2–8 minutes each) covering analog/digital signal processing (9 terms) and semiconductor manufacturing with optics/communication (27 terms), featuring animated diagrams, process visualizations, and real-world examples consistent with multimedia learning principles. Students viewed each video at least twice with in-class checks. The traditional condition involved teacher-led instruction with board explanations and textbook definitions, using no multimedia materials. To maintain content alignment, the same vocabulary items, lesson objectives, and time allocation were used across both formats. Stage 1 (Weeks 2–6) introduced nine foundational signal-processing terms; Stage 2 (Weeks 6–10) addressed 27 conceptually dense terms. Group A received video-based instruction in Stage 1 and traditional in Stage 2; Group B received the reverse. The vocabulary sets were pedagogically parallel but not equivalent in difficulty; the arrangement is not treated as a crossover experiment.

The ten-week study timeline is detailed in Table 2. The two-week retention interval was based on established memory research (Bahrick, 1984; Barclay, 2021; DeHaan et al., 2010). No washout period was implemented between stages, reflecting authentic curriculum progression. Vocabulary tests were administered at five time points (pre-test, immediate post-tests, delayed post-tests), using multiple formats (multiple-choice, fill-in-the-blank, translation, short answer) to capture receptive and productive knowledge. Piloting yielded Cronbach's alpha exceeding .80. The Engagement Questionnaire included Likert-scale items

(perceived usefulness, clarity, engagement, learning efficiency, motivation) and open-ended questions. Semi-structured interviews (n =16, in Mandarin Chinese, 10–15 minutes each) were transcribed verbatim and translated by the researcher. Instruments are provided in Appendices 1–3.

TABLE 2
Ten-Week Instructional Timeline

Week(s)	Group A	Group B	Activities
Week 1	No instruction	No instruction	Pre-test
Weeks 2-3	Video-based (Set 1-2)	Traditional (Set 1-2)	Stage 1 content delivery
Week 4	Review	Review	Immediate Post-Test 1
Week 5	No new content	No new content	Retention interval
Weeks 6-7	Traditional (Set 3-4)	Video-based (Set 3-4)	Stage 2 content delivery
Week 8	Review	Review	Immediate Post-Test 2
Week 9	No new content	No new content	Retention interval
Week 10	-	-	Delayed Post-Test 2; Questionnaire; Interviews

3. Data Analysis, Validity, Reliability and Ethical Considerations

Consistent with the intact-class design, analysis focused on within-group patterns rather than between-group statistical comparisons. Inferential statistics were not used because the non-equivalent groups and non-equivalent vocabulary sets across stages violated the assumptions required for such tests; within-group descriptive comparisons were therefore the only appropriate analytical approach. Quantitative data were entered into SPSS, with descriptive statistics (*M, SD*s) calculated for each group at each assessment point. The pre-test served a descriptive function only—confirming non-equivalence—with no gain scores or inferential analysis. Immediate post-tests were interpreted as criterion-referenced indicators of acquisition; retention was operationalised as descriptive comparison between immediate and delayed means within each group and stage, with no t-tests applied. No statistical comparisons were made between groups, and no causal claims are advanced. Qualitative data from open-ended responses and interviews were analysed using thematic analysis (repeated reading, initial coding, refinement into categories, development of themes), attending to within-group perspectives. Quantitative and qualitative strands were integrated guided by convergence, complementarity, and explanatory depth, with qualitative findings contextualising within-group trends.

Content validity was ensured by aligning test items with taught vocabulary. Construct validity was strengthened through multiple item formats (receptive and productive) and paired immediate-delayed post-tests. Baseline non-equivalence was addressed by refraining from between-group comparisons and interpreting findings strictly within groups.

Procedural validity was supported through standardised testing conditions and matched instructional duration. Reliability was established via pilot testing (Cronbach's α > .80) and consistent scoring procedures. Trustworthiness was enhanced through methodological triangulation (questionnaires and interviews), Mandarin Chinese interviews for nuanced expression, verbatim transcription, consistent protocols, and analytic memos. Transferability was supported through detailed contextual descriptions. Mixed-methods validity was ensured through systematic integration guided by convergence and complementarity.

The study received institutional approval. Students were informed of the study purpose and their rights; participation was voluntary, written informed consent was obtained, and withdrawal was permitted without penalty. Confidentiality was maintained through numerical identifiers, with no names appearing in the dataset; electronic files were password-protected. Fairness was ensured as both instructional conditions represented normal teaching practices, and no research assessment data were used toward course grades.

IV. RESULTS

1. Quantitative Results

Data were collected from two intact ESP classes: Group A (n = 41) and Group B (n = 25). Due to normal attendance fluctuations, valid responses varied slightly across instruments. Specifically, five students from Group B were absent on the pre-test day due to scheduling conflicts, resulting in a valid pre-test sample of 20 for Group B. All 25 students participated in subsequent assessments. All analyses were conducted on a within-group basis; no statistical comparisons were made between groups or across stages. The pre-test contained 25 items (one point per item), for a maximum possible score of 25. Stage 1 post-tests (immediate and delayed) contained 15 items each, for a maximum of 15. Stage 2 post-tests contained 17 items each, for a maximum of 17.

Pre-test results (Table 3) confirmed baseline non-equivalence: Group A (M = 13.60, SD = 4.22) outperformed Group B (M = 7.08, SD = 5.16). For immediate post-tests (Tables 4-5), Group A performed higher in Stage 1 (video-based, M = 12.10, SD = 2.41) than Stage 2 (traditional, M = 8.86, SD = 2.69); Group B showed similar M across Stage 1 (traditional, M = 6.70, SD = 4.90) and Stage 2 (video-based, M = 6.10, SD = 4.20), with wide dispersion. Delayed post-tests (Tables 6-7) showed retention declines: Group A declined modestly (Stage 1: 12.10→11.20, SD = 4.21; Stage 2: 8.86→8.25, SD = 3.53); Group B showed larger declines (Stage 1: 6.70→4.60, SD = 4.60; Stage 2: 6.10→3.30, SD = 3.20), with increased variability.

TABLE 3
Pre-Test Performance

Group	*n*	*M*	*SD*
Group A	41	13.60	4.22
Group B	20	7.08	5.16

Note: Group B originally had 25 participants; five were absent on the pre-test day due to sc heduling conflicts.*

TABLE 4
Immediate Post-Test — Group A

Stage	Mode	*n*	*M*	*SD*
Stage 1	Video-based	33	12.10	2.41
Stage 2	Traditional	40	8.86	2.69

TABLE 5
Immediate Post-Test — Group B

Stage	Mode	*n*	*M*	*SD*
Stage 1	Traditional	25	6.70	4.90
Stage 2	Video-based	25	6.10	4.20

TABLE 6
Delayed Post-Test — Group A

Stage	Mode	*n*	*M*	*SD*
Stage 1	Video-based	35	11.20	4.21
Stage 2	Traditional	34	8.25	3.53

TABLE 7
Delayed Post-Test — Group B

Stage	Mode	*n*	*M*	*SD*
Stage 1	Traditional	24	4.60	4.60
Stage 2	Video-based	24	3.30	3.20

Likert-scale questionnaire results (Table 8, $N = 61$) indicated positive perceptions of video-based instruction. Both groups rated perceived usefulness highly (Group A $M = 4.03$, Group B $M = 4.55$), relevance to major (3.74, 4.32), self-reported motiva tion (3.64, 4.36), perceived engagement (3.56, 4.27), and learning efficiency (3.79, 4.55). The lowest-rated item was “I can remember vocabulary after watching videos” (3.15, 3.73), aligning with retention findings. Both groups strongly supported contin ued video use (4.00, 4.64).

TABLE 8
***M* and *SD*s for Likert-Scale Items (1–5)**

Item	Group A *M* (*SD*)	Group B *M* (*SD*)
1. Videos help me understand ESP vocabulary	4.03 (0.69)	4.55 (0.67)
2. Content is relevant to my major	3.74 (0.80)	4.32 (0.65)
3. Videos increase my motivation	3.64 (0.77)	4.36 (0.66)
4. Videos are more interesting than traditional	3.51 (0.78)	4.23 (0.75)
5. Video length and pacing are appropriate	3.18 (0.77)	3.27 (0.63)
6. I can remember vocabulary after watching	3.15 (0.90)	3.73 (0.83)
7. Videos increase my engagement	3.56 (0.74)	4.27 (0.70)
8. Videos help me complete class tasks	3.67 (0.78)	4.36 (0.66)
9. I want ESP courses to continue using videos	4.00 (0.74)	4.64 (0.58)
10. Videos improve my learning efficiency	3.79 (0.77)	4.55 (0.67)

2. Qualitative Findings and Integration

Open-ended questionnaire responses from 39 Group A and 22 Group B students yielded five recurrent themes. Theme 1 (Perceived advantages): Students emphasised flexibility, repeatability, and clarity, valuing the ability to pause ("随时暂停") and replay ("反复观看/回放"), noting that animations helped them "读懂英文句子" and grasp technical concepts more easily. Theme 2 (Difficulties and challenges): Some noted attention difficulties ("注意力难以持续"), comprehension issues ("有些看不懂"), or pacing mismatches. Theme 3 (Comparison with traditional teaching): Most viewed videos as more vivid and reviewable than traditional "one-time" explanations; several noted the two approaches "各有千秋" (each has strengths). Theme 4 (Suggestions for improvement): Learners proposed shorter videos, clearer vocabulary emphasis, and stronger classroom integration. Theme 5 (Impact on motivation and confidence): Many reported feeling "更放松，更自由" (more relaxed and free) and more willing to review independently, though benefits were not universal.

Semi-structured interviews with 16 students (8 per group) revealed five themes with frequency noted. Theme 1 (Enhanced comprehension through visuals; 13/16): Learners described videos as "更好懂" (easier to understand), attributing clarity to "场景," "实物," and "演示" (scenes, real objects, demonstrations). A3 stated, "比书本干看好懂" (easier than reading the textbook), aligning with Dual Coding Theory. Theme 2 (Autonomy and flexibility; 11/16): Students valued learning during "碎片时间" (fragmented time) and the ability to pause and replay. A2 noted, "灵活，碎片时间能学，短时长也不容易走神" (flexible, helps me stay focused), reflecting Cognitive Load Theory's learner-controlled pacing. Theme 3 (Motivation and confidence; 10/16 positive, 4 neutral, 2 no change): Students reported intrinsic curiosity ("想看看下一个视频讲啥"—curious about the next video, B3). Learners attributed their increased motivation primarily to two features: the ability to control pacing (pause and replay) which reduced anxiety, and the visual clarity of

animations which made abstract processes more concrete. Theme 4 (Challenges; 9/16): Students cited insufficiently deep explanations ("有些术语解释不够深," A6), attention lapses ("注意力不集中," A4), and lack of interaction ("缺少了互动," B7), indicating perceived cognitive overload. Cognitive load occurred most frequently during Stage 2 videos covering semiconductor manufacturing, where learners described terminology as "too dense" and "difficult to follow" even after replaying, suggesting that high intrinsic load from unfamiliar technical systems was not fully mitigated by video features alone. Theme 5 (Recommendations for improvement; 14/16): Suggestions included interactive elements, clearer term explanations, and shorter videos. B4 summarised, "希望多加点互动练习，术语解释再深入些" (more interactive exercises, deeper term explanations).

The mixed-methods integration revealed a broadly convergent pattern: learners perceived video-based instruction as clear and engaging, particularly for concrete, visually representable items—aligning with stable performance patterns. However, in Stage 2 (conceptually demanding vocabulary), performance was lower and more variable across both instructional modes, yet many learners still reported that videos reduced anxiety and supported comprehension even when measurable gains were modest, suggesting that perceived support and performance outcomes do not always align directly. Learners appreciated video-based input while also valuing teacher explanation and guided practice, indicating that video-based instruction functioned most effectively as a complementary rather than stand-alone approach. Overall, video-based instruction is a promising but context-dependent option for ESP vocabulary learning, most useful for foundational technical vocabulary, with outcomes shaped by vocabulary complexity, baseline proficiency, and learner preferences.

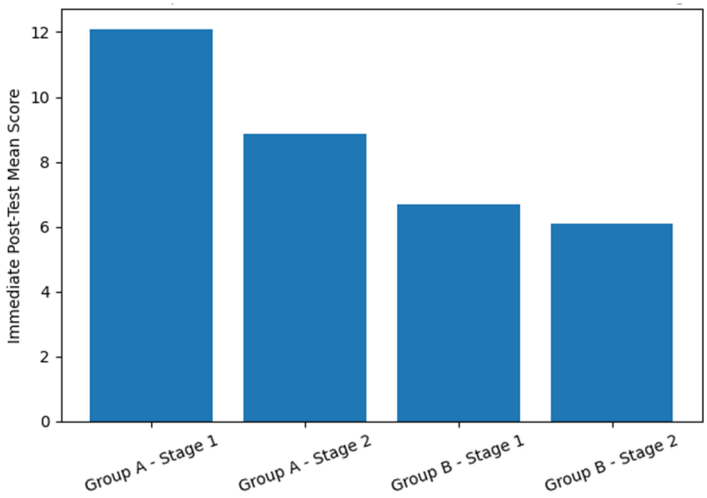
V. DISCUSSION AND IMPLICATIONS

1. Interpretation of Findings in Relation to Research Questions

Research Question 1 (ESP Vocabulary Acquisition): Video-based instruction was associated with stronger performance for foundational, visually representable terminology but weaker outcomes for conceptually dense vocabulary. As shown in Figure 3, Group A (video-based in Stage 1) scored 12.10 on foundational terms vs. 8.86 on complex Stage 2 terms (traditional). Group B scored 6.70 (traditional, Stage 1) and 6.10 (video-based, Stage 2). Qualitative reports indicated animations and diagrams aided visualisation and recall, consistent with Dual Coding and Multimedia Learning Theories. However, Stage 2 performance was lower and more variable, particularly for Group B, with interview comments describing videos as "fast," "dense," or "difficult to follow"—suggesting video

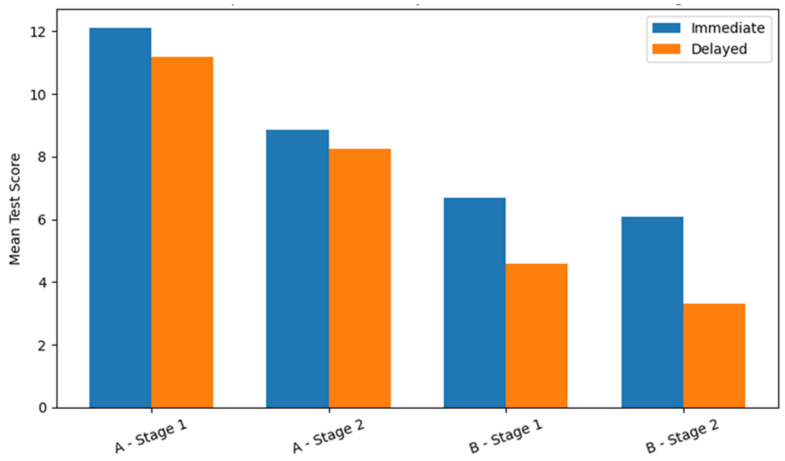
alone may not sufficiently scaffold abstract ESP content. From a Cognitive Load Theory perspective, when intrinsic load is high and prior knowledge limited, multimodal benefits may be offset by processing demands. The relationship is conditional, shaped by vocabulary complexity, learner proficiency, and cognitive demands.

FIGURE 3
Within-Group Immediate Post-Test Means Across Instructional Stages



Research Question 2 (ESP Vocabulary Retention): As shown in Figure 4, retention was stronger for foundational Stage 1 vocabulary than for conceptually denser Stage 2 terminology. For Stage 1 (concrete, visually representable terms), both groups showed modest decline: Group A dropped from 12.10 to 11.20; Group B from 6.70 to 4.60. Qualitative reports indicated visual demonstrations enabled learners to "picture the process" when recalling terminology, consistent with dual-coding and multimedia learning perspectives. For Stage 2 (abstract, procedurally dense terms), a more constrained pattern emerged: Group A dropped from 8.86 to 8.25; Group B from 6.10 to 3.30, with greater variability. Learners described Stage 2 content as dense, fast-paced, or difficult to understand even with video support, aligning with Cognitive Load Theory: when intrinsic load is high and prior knowledge limited, new vocabulary is harder to consolidate and more vulnerable to forgetting. Video-based instruction is valuable for reinforcing accessible technical vocabulary, but requires complementary teacher mediation for highly specialised or abstract terminology.

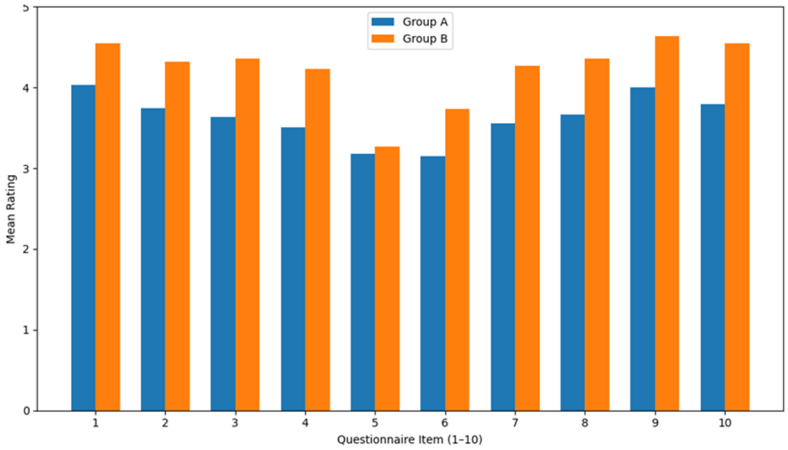
FIGURE 4
Within-Group Immediate and Delayed Post-Test Means Across Stages



Research Question 3 (Learner Engagement and Perceptions): As shown in Figure 5, mean ratings across the ten Likert-scale items were above the scale midpoint for both groups, indicating generally positive perceptions. Group B reported slightly higher ratings than Group A, interpreted descriptively given baseline differences. Learners viewed videos as useful, engaging, efficient, and relevant, with strong endorsement for continued use (Group A $M = 4.00$; Group B $M = 4.64$). Learners stated that videos made ESP vocabulary easier to understand, particularly with visual demonstrations and step-by-step explanations, described as "intuitive," "clear," or "less abstract." Motivational responses were also positive: videos were perceived as "refreshing" and "easy to focus on." A prominent theme concerned autonomy: students valued pausing and replaying content, associating these features with reduced anxiety and increased confidence (Ryan & Deci, 2000). Learners attributed their increased motivation primarily to two features: the ability to control pacing (pause and replay) which reduced anxiety, and the visual clarity of animations which made abstract processes more concrete. Despite positive views, learners also identified challenges, including pacing that felt "too fast" or "too dense," difficulty processing advanced terms, and limited interaction. This lack of interaction reflects the asynchronous nature of video-based instruction, which provides no real-time teacher feedback or peer discussion. Cognitive load occurred most frequently during Stage 2 videos covering semiconductor manufacturing, where learners described terminology as "too dense" and "difficult to follow" even after replaying, suggesting that high intrinsic load from unfamiliar technical systems was not fully mitigated by video features alone. These reservations mirror the less stable performance for complex vocabulary in Stage 2. Many preferred combining video-based instruction with

teacher-led explanation. These results indicate that video-based instruction is well-received as a supportive component when implemented as part of a blended, scaffolded model rather than as a standalone method.

FIGURE 5
Mean Engagement Questionnaire Ratings by Group (Items 1–10)



2. Implications (Pedagogical, Theoretical, Methodological)

Pedagogically, short (3–7 minutes), clearly structured videos with visual diagrams help learners manage cognitive load for foundational vocabulary. Classroom tasks (guided note-taking, vocabulary mapping) should build on this exposure. Embedding videos at multiple points in the learning cycle (review, flipped preparation, pre-assessment revision) supports distributed reinforcement. Videos accommodate learner diversity through pausing, replaying, and self-pacing, reducing anxiety and increasing confidence. However, video alone is insufficient for highly technical terminology; teacher-mediated scaffolding (clarification, modelling, bilingual glossing, guided practice) remains essential. Design refinements include clear signposting, slower pacing, concrete examples, authentic workplace links, and interactive elements. Theoretically, the findings suggest a lexical–conceptual threshold effect (Laufer, 2013): durable acquisition requires a minimum level of conceptual interpretability. The study confirms that visual–verbal integration supports meaning-making but highlights boundary conditions—segmentation alone cannot fully mitigate high intrinsic load from conceptually dense terminology. A distinction emerges between perceived cognitive fluency and actual conceptual mastery. The study also contributes to Self-Determination Theory by showing how autonomy-supportive features (pacing control, replay options) reduce anxiety and support self-regulation. ESP learning emerges as an

ecology of interacting supports—multimedia input, teacher mediation, and applied disciplinary practice playing complementary roles. Methodologically, the study demonstrates the value of parallel-group orientation with within-group, stage-specific analyses in intact classrooms where random assignment is infeasible. Mixed-methods integration clarified why positive attitudes did not always correspond to measured performance. The multi-modal assessment strategy (recognition, fill-in-the-blank, translation, short answer) captured both receptive and productive dimensions. Substantial individual variation suggests future research should adopt fine-grained analytical approaches (learner profiling, longitudinal case analysis).

3. Limitations and Directions for Future Research

Several limitations should be acknowledged: single institution with two intact classes, limiting generalisability; lack of random assignment resulting in baseline proficiency differences; ten-week duration capturing only short-term retention; traditional instruction reflecting instructor style and contextual factors; non-standardised researcher-developed tests; self-reported engagement measures subject to bias; unmeasured individual learner variables (digital literacy, self-regulation, prior knowledge, motivation); the study conceptualized engagement holistically without distinguishing behavioral, cognitive, and emotional dimensions; and motivation/engagement were measured using single Likert-scale items, which may not capture their multi-dimensional nature. Building on these, future research should extend to larger, more diverse populations across multiple institutions and disciplines. Longitudinal designs are needed to examine sustained retention and transfer to workplace communication. Research should investigate pedagogical sequencing effects (pre-class, in-class, post-class positioning) within blended or flipped models. Design optimisation studies should examine video length, segmentation, pacing, interactive prompts, embedded quizzes, and bilingual glosses. Individual learner differences should be addressed through profiling studies or person-centred analyses, and future studies should also examine each engagement dimension (behavioral, cognitive, emotional) separately rather than holistically. Researcher-developed measures should be combined with standardised ESP assessments and authentic performance tasks (interpreting technical manuals, producing workplace documentation). Finally, incorporating teachers' perspectives through interviews and design-based research would illuminate workload, curriculum integration, and professional-development requirements for sustainable adoption.

VI. CONCLUSION

This study examined video-based instruction for ESP vocabulary learning among Chinese vocational university students using a mixed-methods parallel-group quasi-experimental design over ten weeks. Three key insights emerge: ESP vocabulary learning is concept-dependent, with learners succeeding most on concrete, visually representable terms where lexical form maps onto technical systems; learning abstract or conceptually dense terminology is constrained by cognitive load and prior knowledge, with multimodal presentation facilitating learning only when processing demands remain manageable; and the perceived motivational benefits of video-based instruction—autonomy, controllable pacing, repeated exposure—may contribute to an affective environment that supports persistence and reduces emotional barriers.

The study contributes to ESP vocabulary theory by demonstrating that meaning-making depends on conceptual systems underpinning terminology; to cognitive-linguistic perspectives by showing that dual coding is salient when language maps onto technical processes and that cognitive load principles clarify when multimodal input is facilitative; to motivation research by illustrating how pace control and repeatability provide affective support; and methodologically by demonstrating rigorous, ecologically valid research in intact classes through within-group interpretation and mixed-methods integration. Pedagogically, video-based instruction plays a constructive role within blended models for foundational vocabulary when sequenced across the learning cycle (pre-class, in-class, review) and followed by teacher-led clarification, though its value is contingent—outcomes varied for conceptually dense terminology, indicating prior knowledge and scaffolding remain critical.

These findings align with previous research showing that video-based instruction supports vocabulary acquisition for concrete, visually representable terms (Peters & Muñoz, 2020; Plass & Jones, 2005), but extend it by revealing that conceptually dense terminology requires additional teacher scaffolding—a boundary condition not fully addressed in prior studies. Accordingly, specific pedagogical guidelines for ESP practitioners include: (a) use short videos (3–7 minutes) with clear visual support for foundational technical vocabulary; (b) embed interactive elements (e.g., pause-and-answer prompts, comprehension checks) to address interaction limitations, especially for complex content; (c) sequence videos across the learning cycle (pre-class preview, in-class reinforcement, post-class review); and (d) combine video-based instruction with teacher-led clarification and guided practice in a blended, scaffolded model. Limitations include a single institution, two intact classes, within-group interpretations, researcher-developed assessments, self-reported data, and single-item measures of motivation and engagement.

Future research should examine the link between lexical form and conceptualisation using

discourse-analytic or eye-tracking methods, determine cognitive load thresholds, profile learner variability, examine transfer to task-based performance, conduct longitudinal studies, and incorporate teacher-focused inquiry. When thoughtfully integrated into blended teaching, video-based instruction can broaden access to specialised language learning, support learner autonomy, and strengthen links between language and professional practice.

REFERENCES

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, *113*(1), 1–29. https://doi.org/10.1037/0096-3445.113.1.1

Coxhead, A. (2025). Vocabulary and English for specific purposes. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (2nd ed., pp. 107–123). Wiley Blackwell. https://doi.org/10.1002/9781394264726.ch6

Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior*, *89*, 465–470. https://doi.org/10.1016/j.chb.2018.07.015

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*,*74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Ghafar, Z., Abdulkarim, S. T., Mhamad, L. M., Kareem, R. A., Rasul, P. A., & Mahmud, T. I. (2023). Microlearning as a learning tool for teaching and learning in acquiring language: Applications, advantages, and influences on the language. *Canadian Journal of Educational and Social Studies*, *3*(2), 45–62. https://doi.org/10.53103/cjess.v3i2.123

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the First ACM Conference on Learning@Scale Conference* (pp. 41–50). https://doi.org/10.1145/2556325.2566239

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*(3), 539–558. https://doi.org/10.1111/0023-8333.00164

Lockhart, R. S., & Craik, F. I. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology/RevueCanadienne de Psychologie, 44*(1), 87–112. https://doi.org/10.1037/h0084237

Mayer, R. E. (2002). Multimedia learning. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 41, pp. 85–139). Academic Press. https://doi.org/10.1016/S0079-7421(02)80005-6

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/CBO9780511816819.004

Nation, P. (2017). How vocabulary is learned. Indonesian JELT: *Indonesian Journal of English Language Teaching, 12*(1), 1–14. https://doi.org/10.25170/ijelt.v12i1.145

Paivio, A., & Clark, J. M. (1991). Dual coding theory and education. *Educational Psychology Review, 1*, 149–210.

Radvansky, G. A., Doolen, A. C., Pettijohn, K. A., & Ritchey, M. (2022). A newlook at memory retention and forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(11), 1698–1723. https://doi.org/10.1037/xlm0001086

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484–503. https://doi.org/10.1017/S0261444812000018

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (2011). Cognitive load theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation* (Vol. 55, pp. 37–76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Examples in: English
Applicable Language: English
Applicable Levels: Tertiary

Xingxue Liu (instructor)
Shandong Vocational University of Foreign Affairs
School of International Exchange
No.788, Changjiang Rd, Rushan, Weihai, Shandong, China
Email: 905726744@qq.com

Sunmee Chang (professor)
Dept. of English Language and Literature

Hoseo University
12, Hoseodae-gil Dongnam-gu
Cheonan, Chungnam, South Korea
Email: schang@hoseo.edu

Received: April 29, 2026
Reviewed: May 30, 2026
Revised version received: June 25,2026