MULTIMEDIA-ENHANCED ESP INSTRUCTION IN HIGHER EDUCATION: INTEGRATING INTERACTIVE VIDEO, PROFESSIONAL SIMULATIONS AND GENERATIVE AI

Dilfuza Sh. Shavkieva

PhD, Associate Professor, Gulistan State University

Gulistan, Republic of Uzbekistan

Abstract. This conceptual and design-oriented article examines the pedagogical integration of multimedia technologies into English for Specific Purposes (ESP) instruction at universities. Drawing on needs analysis, multimedia learning theory, task-based language teaching, and recent research on generative artificial intelligence, the study proposes a five-stage instructional cycle combining authentic professional video, multimodal scaffolding, simulation, AI-supported rehearsal, and performance-based assessment. An illustrative module for scientific and workplace communication demonstrates how language objectives can be aligned with realistic disciplinary tasks. The analysis differentiates established research findings from proposed instructional design and identifies practical constraints involving cognitive load, feedback reliability, data protection, accessibility, and teacher preparation. No classroom experiment or numerical gains are claimed. The model offers a transferable framework for cautious, measurable implementation in non-philological higher education.

Keywords: English for Specific Purposes, ESP, multimedia learning, interactive video, task-based instruction, generative AI, professional communication, multimodal pedagogy.

INTRODUCTION

The teaching of English for Specific Purposes increasingly requires more than acquisition of terminology or grammatical accuracy. University learners must interpret technical information, explain procedures, collaborate across professional roles and communicate through digital channels. In medicine, engineering, biology, economics and education, professional interaction routinely combines speech, written documents, diagrams, screen-based demonstrations and visual data. Language pedagogy that treats these modes separately may fail to prepare learners for the communicative ecology of contemporary workplaces.

A multimedia resource is pedagogically useful not because it contains audio or animation but because it supports a defined language-performance outcome. A scientific video may require learners to summarize evidence; a simulated client meeting may require clarification and negotiation; an annotated chart may require explanation of trends and limitations. When such tasks are sequenced around professional needs, ESP becomes a site for integrating disciplinary meaning and communicative competence. This integration is particularly significant in non-philological university programmes where contact hours for language instruction may be restricted.

Recent progress in accessible video production, interactive captions, speech recognition and generative AI has expanded the range of feasible ESP activities. Yet adoption also creates a design problem: attractive media can overload learners, automated feedback may be misleading, and scripted AI exchanges may not reproduce the social demands of actual professional encounters. Therefore, technology choice should follow linguistic objectives, not determine them.

The purpose of this article is to develop and critically examine a multimedia-enhanced ESP instructional model for higher education. Its central research question is: how can multimodal input, authentic tasks and AI-supported practice

be coordinated while preserving disciplinary accuracy, learner agency and transparent assessment? The article presents a conceptual synthesis and an illustrative teaching design rather than reporting an intervention conducted with students.

LITERATURE REVIEW

ESP scholarship has long argued that instructional content should be determined by the demands of target situations. Hutchinson and Waters (1987) conceptualize ESP as a learning-centred enterprise, while Dudley-Evans and St John (1998) foreground the relationship between language use, specialized genres and learners' disciplinary purposes. Basturkmen (2010) similarly shows that syllabus design requires decisions about needs analysis, language description, material selection and evaluation. These approaches caution against equating ESP with lists of technical terms: the target is communicative participation in a professional community.

Task-based perspectives contribute an additional principle. Ellis (2003) explains how meaningful tasks can organize opportunities for language use, while Long (2015) emphasizes real-world target tasks and principled pedagogical sequencing. An ESP activity should consequently produce a communicative outcome, such as an oral safety briefing, a consultation summary or a research poster explanation, rather than simply ask students to repeat expressions presented in a video.

The cognitive theory of multimedia learning suggests that learners process visual and verbal information through partially distinct channels with limited capacity. Mayer (2021) stresses the instructional significance of selecting relevant material, managing essential processing and supporting generative understanding. In ESP settings, this perspective supports segmenting technical explanations, synchronizing diagrams with narration, and avoiding decorative visuals that compete with specialized terminology. A multimedia-rich lesson is not automatically an effective lesson.

Research into multimodal immersion in higher education offers additional support for considering communication as an orchestration of linguistic, visual and embodied resources. A 2024 systematic review of thirty-four studies reported the educational relevance of multimodal approaches for English learning in academic and specific-purpose contexts, while also showing variation in methods and learning environments. These findings justify continued exploration but do not establish that every digital medium is equally beneficial for every learner or discipline.

Generative AI has introduced further possibilities for adaptive dialogue, draft revision and low-stakes practice. Li et al. (2025), reviewing 144 peer-reviewed empirical articles published in 2023–2024, identify expanding research on teaching strategies, learner perceptions, writing, feedback and implementation. The review also reports that research attention has been distributed unevenly across skills and educational levels. Therefore, conclusions about AI-based improvement in professional speaking or long-term transfer require particular caution.

UNESCO's guidance on generative AI in education and research (Miao and Holmes, 2023) insists on human-centred safeguards and identifies privacy, governance and educational validation as major concerns. An AI conversation partner can create practice opportunities, but its responses should not be treated as authoritative professional advice. In specialized English courses, disciplinary content must remain open to verification against reliable documents and teacher or expert judgment.

Together, these literatures suggest a coherent instructional principle: connect needs analysis with authentic tasks; use multimedia to make complex professional communication comprehensible; provide opportunities to perform and reflect; and apply AI only where it supports these aims with defensible safeguards. The resulting model should be judged by the quality of the communicative performance it enables, rather than by the number of digital applications used.

MATERIALS AND METHODOLOGY

This article uses qualitative conceptual synthesis and pedagogical design analysis. Its materials consist of published ESP methodology, multimedia learning research, and recent reviews and guidance concerning generative AI and multimodal English teaching. The discussion treats these publications as conceptual and empirical foundations with different evidential status. It does not claim a systematic review, an independently collected learner corpus, classroom observations or an experimental dataset.

The model was constructed through four analytical steps. First, target communicative events were identified at the level of representative professional activity: explaining a procedure, interpreting evidence, reporting a problem and negotiating a response. Second, each event was mapped onto language functions such as sequencing, hedging, clarification, comparison and confirmation. Third, appropriate digital affordances were selected, including replayable video, synchronized captions, annotated visual data, simulation and optional AI rehearsal. Fourth, a performance rubric was specified to make the educational outcomes observable.

A practical design assumption is that the target group consists of university students with an intermediate or mixed English proficiency profile. The proposed module is adjustable for different disciplines and is illustrated here through scientific workplace communication. All sample tasks and feedback prompts are pedagogical constructions. They should be piloted with actual learners, calibrated to local professional requirements and evaluated before claims of effectiveness are made.

The evaluation logic distinguishes feasibility from effectiveness. Feasibility concerns whether students can access and complete activities, understand instructions and receive usable feedback. Effectiveness would require independent evidence of learning, preferably through pre-specified performance assessments, comparable tasks, blinded scoring where possible and appropriate consideration of baseline proficiency. The present contribution addresses the first design stage and outlines the evidence needed for subsequent validation.

RESULTS AND PEDAGOGICAL ANALYSIS

The principal result of this design inquiry is a five-stage multimedia-enhanced ESP cycle: (1) professional needs and task orientation, (2) multimodal input and noticing, (3) guided interaction, (4) authentic performance with optional AI rehearsal, and (5) reflective assessment and transfer. The stages are recursive rather than a rigid software workflow. A teacher can repeat a stage when learners need additional scaffolding or when the disciplinary task proves unexpectedly complex.

Stage 1: professional needs and task orientation. Each instructional unit begins with a plausible communicative problem, such as presenting a laboratory finding to a non-specialist colleague. The teacher identifies the audience, purpose, medium, expected output and language demands. A short baseline attempt reveals prior competence and enables differential scaffolding. This orientation is essential because learners should recognize why particular forms and strategies matter in their field.

Stage 2: multimodal input and noticing. Students watch a brief authentic or carefully adapted professional video with relevant charts, labels or subtitles. The teacher selects a short segment rather than requiring passive viewing of an entire lecture. On the first viewing, students identify the situation and central message; on the second, they note expressions that introduce evidence, indicate uncertainty or signal cause and effect. Captions can initially support access to difficult language and subsequently be reduced to encourage listening. An accompanying transcript permits review without treating subtitles as a substitute for comprehension.

In designing this stage, visual economy matters. Technical figures should be legible and directly connected to the accompanying speech. Highlighting one variable at a time can reduce unnecessary search across a complex diagram. Excessive screen annotations, background music and simultaneous text blocks may divide attention. Learners also need

alternatives when video is inaccessible because of bandwidth, hearing-related needs or visual limitations; transcripts, downloadable audio and text descriptions should form part of the original plan.

Stage 3: guided interaction. Learners work with a partially completed professional communicative task. For example, one student explains a chart and another asks for clarification about units, limitations and practical significance. A language support sheet supplies functional phrases such as 'The data suggest...', 'Could you clarify whether...' and 'One limitation is...'. Rather than memorizing scripts, students use these resources to communicate distinct information. Rotating participant roles promotes listening, reformulation and a more balanced distribution of speaking opportunities.

Stage 4: authentic performance and AI-supported rehearsal. Students prepare a two-minute technical briefing or a five-minute problem-solving consultation. They may rehearse with a conversational AI tool that plays a defined stakeholder role and asks follow-up questions. Prompts should delimit the topic and request diagnostic feedback on clarity, organization and communicative appropriateness rather than an unqualified numerical score. Students must check content claims and may reject AI suggestions that distort disciplinary meaning or produce unnatural professional language.

One illustrative prompt is: 'Act as a colleague who needs a clear explanation of a laboratory result. Ask one question at a time about the trend, possible causes and limitations. Do not invent measurements; if information is missing, ask for it. After the exchange, identify two strengths and two communication improvements.' This prompt makes the simulation transparent while restricting unsupported factual elaboration. Its performance and reliability still require teacher review.

Stage 5: reflective assessment and transfer. Learners submit a recorded briefing, a short written summary and a reflection explaining how feedback influenced their revision. The teacher evaluates the final performance with a consistent rubric. Suggested categories are task completion, accuracy of disciplinary meaning, discourse organization, interactional effectiveness, linguistic range and accuracy, and appropriate multimodal presentation. Each criterion should have explicit descriptors and be weighted according to the actual target situation.

For a laboratory communication module, the first lesson can use a short demonstration video followed by an annotated process diagram. The second lesson can concentrate on a chart-based evidence briefing and paired clarification exchanges. The third can introduce a simulated problem report using teacher-developed role cards. A final task might ask students to report an unexpected result to an international research collaborator, separating observation from interpretation and identifying what remains uncertain. These are suggested activities, not activities documented as implemented at Gulistan State University.

A significant pedagogical advantage of the cycle is its flexible treatment of learner proficiency. Students who need substantial support may receive slower audio, transcripts, lexical glosses and preparation time; more advanced learners can work with incomplete information, competing interpretations and spontaneous questions. Differentiation should alter the amount of support without removing the central communicative purpose. This approach makes technology responsive to learner needs rather than making learners adapt to a fixed application.

Another design implication concerns the distinction between automated correction and professional adequacy. A system may improve surface fluency while overlooking the appropriateness of a scientific hedge or the interpersonal consequences of an instruction. For instance, replacing 'The findings may indicate' with 'The findings prove' could make a statement appear more decisive while making it epistemically less accurate. Therefore, the rubric must explicitly reward calibrated claims and justified uncertainty.

A feasible future pilot could compare two cohorts completing equivalent professional tasks, with both groups receiving the same instructional objectives and contact hours. The multimedia condition would use the five-stage cycle while the comparison condition would employ conventional materials matched for content. Pre- and post-task

recordings could be evaluated with anonymized rubrics by trained raters; access barriers and learner experiences could be investigated through interviews. These proposals describe a research design, not results obtained from participants.

DISCUSSION

The model connects the strengths of ESP needs analysis with contemporary digital affordances, but its validity depends on the realism and quality of the target tasks. A simulation that asks learners to explain a familiar diagram can practice sequencing and clarity; it cannot establish competence in unpredictable workplace situations by itself. Transfer requires opportunities to work with authentic constraints, including incomplete knowledge, different audiences, professional accountability and time pressure.

Three implementation risks require explicit governance. First, cognitive overload may emerge when students are expected to interpret technical content while simultaneously handling a new interface and unfamiliar language. Second, access inequities may disadvantage students with limited connectivity, shared devices or accessibility needs. Third, unreliable generated content can mislead students, especially when technical terms, scientific evidence or professional norms are involved. Teachers need protocols for source verification, data minimization, acceptable tool use and alternatives to AI services.

Teacher expertise remains central. Instructors must decide which communication practices are disciplinary conventions and which represent optional rhetorical choices. Joint preparation with subject specialists can improve authenticity, but pedagogical responsibility cannot be delegated to software. Similarly, peer interaction and direct instructor feedback should remain part of the course. A model built exclusively on individual screen activity could reduce opportunities for negotiation, pragmatic adjustment and collaborative knowledge construction.

Assessment is also an ethical concern. Learners should know whether automated tools are used, what data might be processed and whether recordings are retained. AI feedback should be formative and contestable; consequential grades should rely on transparent criteria and accountable human evaluation. Institutions should provide workable non-AI alternatives when students cannot or do not wish to share personal information with external platforms.

The proposed model's limitations are substantial. It is a conceptual pedagogical framework, not a tested intervention; it presents no participant sample, proficiency gains, inferential statistics or longitudinal transfer evidence. A single illustrative scientific communication module cannot represent the needs of all ESP fields. Future studies should evaluate the model across disciplines and proficiency bands, document teacher workload and technology access, and determine whether benefits persist beyond the immediate instructional setting.

CONCLUSION

Multimedia technologies can strengthen ESP instruction when their selection is guided by professional communication needs and explicit learning outcomes. Interactive video can contextualize disciplinary language, simulations can organize purposeful interaction, and generative AI can provide additional opportunities for rehearsal and revision. None of these affordances, however, guarantees learning without appropriate scaffolding, meaningful tasks and trustworthy assessment.

The five-stage instructional cycle presented in this article integrates needs orientation, multimodal input, guided interaction, performance and reflective transfer. Its main contribution is a transparent design framework that aligns technologies with communicative functions and maintains a clear distinction between proposed pedagogical advantages and empirically demonstrated outcomes. The next research priority is rigorous classroom validation using authentic performance measures, accessible implementation and human oversight of AI-supported feedback.

REFERENCES

Basturkmen, H. (2010). Developing Courses in English for Specific Purposes. Palgrave Macmillan. https://doi.org/10.1057/9780230290518

Dudley-Evans, T., & St John, M. J. (1998). Developments in English for Specific Purposes: A Multi-Disciplinary Approach. Cambridge University Press.

Ellis, R. (2003). Task-Based Language Learning and Teaching. Oxford University Press.

Hutchinson, T., & Waters, A. (1987). English for Specific Purposes: A Learning-centred Approach. Cambridge University Press.

Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. Computers and Education: Artificial Intelligence, 9, 100445. https://doi.org/10.1016/j.caeai.2025.100445

Long, M. (2015). Second Language Acquisition and Task-Based Language Teaching. Wiley Blackwell.

Mayer, R. E. (2021). Multimedia Learning (3rd ed.). Cambridge University Press.

Miao, F., & Holmes, W. (2023). Guidance for Generative AI in Education and Research. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693

Multimodal immersion in English language learning in higher education: A systematic review. (2024). Heliyon. https://doi.org/10.1016/j.heliyon.2024.e38357

Uygun, E., & Demir, B. (2026). A PRISMA-based systematic review of artificial intelligence in English as a foreign and second language education (2023–2025). Discover Education, 5, Article 478. https://doi.org/10.1007/s44217-026-01504-y

Reviewer:
Koumar J. Makesh
Ph.D.
University of Delhi (India)