Journal of Literature, Languages and Linguistics
ISSN 2422-8435 An International Peer-reviewed Journal
Vol.110, 2026

Enhancing Medical Vocabulary Acquisition through Multisensory Integration

Hien Thi Thanh Dao*
Department of Foreign Languages, University of Medicine and Pharmacy at Ho Chi Minh City,
217 Hong Bang Street, Cho Lon Ward, Ho Chi Minh City, Vietnam
* E-mail of the corresponding author: dtthiendhyd@ump.edu.vn

Abstract

Medical terminology is often regarded as one of the most challenging components of English for Specific Purposes (ESP) curricula in health sciences. The complexity stems from the Latin and Greek roots of many terms, their phonological difficulty, and the necessity for precise spelling and pronunciation in clinical communication. Traditional vocabulary teaching methods, which emphasize rote memorization or isolated exposure, have frequently failed to ensure long-term retention. This article proposes a four-step multisensory framework for learning medical terminology that systematically integrates auditory, visual, articulatory, and kinesthetic modalities. The sequence includes: (1) listening to pronunciation, recognizing the orthographic form, and observing phonetic transcription; (2) listening and visual recognition combined with oral repetition; (3) listening, reading, and writing the term; and (4) listening and writing through dictation or recall. By engaging multiple senses in a structured, repetitive cycle, the model enhances both receptive and productive skills, consolidates memory, and fosters learner confidence. The article situates this framework within the context of ESP teaching in Vietnam and discusses its pedagogical implications for classroom practice, curriculum design, and potential digital applications. Ultimately, it argues that multisensory integration offers an effective pathway to mastering medical vocabulary, ensuring accuracy and fluency essential for future healthcare professionals.

Keywords: Multisensory learning; medical terminology; ESP; vocabulary acquisition; medical English.

DOI: 10.7176/JLLL/110-02

Publication date: March 31st 2026

1. Introduction

Mastering medical terminology is a fundamental requirement for students in medicine and pharmacy. Specialized vocabulary enables learners to read academic literature, communicate accurately in clinical settings, and participate effectively in research activities. However, errors in the pronunciation or spelling of complex terms can lead to misunderstandings, diminish professional credibility, and, in some cases, contribute to clinical mistakes. Therefore, instruction in medical terminology within English for Specific Purposes (ESP) programs must extend beyond simple word recognition to support precise, confident use in authentic professional contexts. This challenge is particularly pronounced for students in countries where English is not the primary language, including Vietnam. Many medical terms originate from Latin and Greek, resulting in unfamiliar spelling conventions and phonological patterns. Consequently, long words with complex stress structures are difficult to pronounce, remember, and use accurately. Traditional teaching methods that emphasize memorization of word lists or passive reading often produce only short-term recognition and do not foster durable retention or active communicative ability. To address these limitations, educators have increasingly turned to multisensory approaches to vocabulary learning. Research in cognitive psychology indicates that memory is strengthened when information is processed through multiple sensory channels. When learners simultaneously see a word, hear its pronunciation, articulate it aloud, and write it, they form richer neural connections that enhance comprehension and recall. Although multisensory learning has been widely investigated in general language education, its systematic application to medical terminology remains limited.

In response, this article proposes a structured four-step multisensory framework specifically designed for medical vocabulary acquisition. The sequence progresses from recognition to production. Initially, learners listen to the term, view its written form, and associate it with phonetic symbols. They then repeat the word aloud while continuing to receive auditory and visual input. In the third stage, learners integrate listening, reading, speaking, and writing to reinforce the term. Finally, they demonstrate mastery by writing the word from memory after

hearing it, without visual support. This gradual progression builds confidence while strengthening both receptive and productive skills. Such a framework is particularly relevant in contexts like Vietnam, where students frequently report difficulty retaining lengthy technical terms and have limited opportunities for active practice beyond the classroom. By engaging multiple senses, vocabulary learning becomes more interactive, meaningful, and effective rather than passive. Moreover, the approach aligns well with contemporary digital tools, many of which integrate audio, text, and interactive writing functions. The remainder of this article reviews relevant literature on multisensory learning, explains the four-step framework in detail, and discusses implications for teachers and curriculum designers in medical English programs. Ultimately, it aims to provide a practical, research-informed strategy for improving the acquisition of medical terminology in non-native English learning environments.

2. Literature Review

2.1 Multisensory Learning and Cognitive Theories

Multisensory learning is an approach that involves using more than one sense at the same time during learning. Instead of relying only on reading or listening, learners see, hear, speak, and sometimes physically interact with the material. Cognitive psychology research shows that this combination helps the brain store and recall information more effectively. When several senses are involved, learners create stronger mental connections, which makes it easier to remember what they have learned later. Paivio's Dual Coding Theory explains that verbal information such as spoken or written words and non-verbal information such as images or sounds are processed through two different but connected systems in the brain. Learning becomes more effective when both systems work together. In vocabulary learning, this means that seeing how a word is written, hearing how it sounds, and producing it through speech or writing all reinforce each other and support long-term memory.

Mayer's Cognitive Theory of Multimedia Learning also supports the value of multisensory input. According to this theory, students learn best when words and images are presented together in a clear and organized way that does not overload the mind. When learners receive information through both auditory and visual channels, they can process it more deeply. In language learning, listening provides sound patterns, reading provides visual form, and speaking or writing adds physical involvement. These channels complement each other and help learners build a more complete understanding of new vocabulary. Together, these cognitive theories provide a strong foundation for using multisensory techniques in teaching specialized language, including medical terminology.

2.2 Multisensory Approaches in General Language Learning

In second language learning, many studies have shown that using multiple senses improves vocabulary acquisition. Learners differ in their preferred learning styles, some learn better through listening, others through visual input, and others through movement or hands-on activity. However, research suggests that combining these modes usually produces the best results for most learners. Another important principle is repeated exposure. Vocabulary is not learned after one encounter; it requires meeting the word many times in different contexts. Multisensory activities naturally provide this repetition because learners see, hear, say, and write the same word in various ways.

Empirical studies confirm that combining auditory and visual input leads to better recall than using only one type of input. For instance, students who both listen to and read new words tend to remember them more accurately than those who only read them. Adding writing tasks strengthens learning even further because learners must actively produce the word rather than simply recognize it. This idea relates to the depth of processing principle, which states that information processed more actively and meaningfully is remembered better. In classroom practice, teachers often use techniques such as dictation, reading aloud, phonetic transcription, and multimedia flashcards that include images, sound, and text. Research reviews indicate that such multisensory training not only improves how well students learn new words but also helps them recall them more quickly, which is important for real communication.

2.3 ESP and Vocabulary Challenges in Medical Education

English for Specific Purposes (ESP), especially medical English, presents additional difficulties beyond general language learning. Medical terminology is typically long, complex, and based largely on Latin and Greek roots. Words such as gastroenterology, hypertension, or bronchopneumonia contain many syllables and unfamiliar structures, making them difficult for learners who are used to everyday vocabulary. Pronunciation problems are common, and spelling errors often remain even among advanced students. Because these terms are essential for professional communication, inaccurate use can affect both academic success and future clinical practice.

Research in ESP has emphasized that specialized vocabulary is not just one part of learning but the foundation of professional competence. Without a strong command of technical terms, students struggle to understand lectures, textbooks, and clinical materials. However, teaching methods frequently depend on memorizing lists, matching words with definitions, or using simple flashcards. While these methods may help students recognize terms temporarily, they often fail to produce lasting mastery or confident use in speaking and writing. Studies involving healthcare students have shown that insufficient knowledge of medical terminology can hinder both academic performance and effective communication in professional settings.

2.4 Multisensory Learning in Medical and Health Professions Education

Multisensory teaching is widely used in medical education, but mostly in areas such as anatomy, physiology, and clinical skills rather than vocabulary learning. Students often learn through visual models, simulations, videos, and hands-on practice, which help them understand complex structures and procedures. For example, using three-dimensional anatomical models together with spoken explanations has been shown to improve retention compared with studying text alone. These methods demonstrate the clear benefits of engaging multiple senses in medical training.

However, when it comes to learning medical terminology itself, research is still limited. Some recent studies suggest that multimodal tools such as audiovisual glossaries or flashcards that include pronunciation can help learners remember technical terms. Despite these promising findings, there is little evidence of a systematic approach that guides students step by step through multisensory vocabulary learning. Most practices are informal or unstructured. This lack of a clear framework highlights the need for models that deliberately combine listening, reading, speaking, and writing to support the acquisition of specialized vocabulary in medical English.

2.5 Research Gap and Rationale for the Study

Overall, the literature points to several important conclusions. First, cognitive theories strongly support the idea that learning becomes more effective when multiple sensory channels are used together. Second, research in general second language acquisition shows clear benefits of combining auditory, visual, and productive activities for vocabulary learning. Third, despite the importance of terminology in ESP, especially in medical education, there is still a shortage of structured multisensory approaches specifically designed for this context.

This study addresses that gap by proposing a four-step multisensory framework tailored to medical vocabulary learning. Unlike isolated or informal multimodal activities, the framework offers a clear progression from recognizing a term to producing it independently. Learners interact with each word through listening, seeing, speaking, and writing in a planned sequence, which strengthens both understanding and retention. The approach is particularly relevant for students in non-native English environments, such as Vietnam, where opportunities to practice medical English outside the classroom may be limited. By grounding the framework in established cognitive theory while responding to real educational needs, the study aims to provide a practical strategy for improving the learning of medical terminology.

3. The Proposed Framework

The proposed Four-Step Multisensory Framework is designed to help students learn medical vocabulary in a gradual and supportive way, moving from simple recognition to confident, independent use. Medical terms are often long, complex, and unfamiliar, so students need more than memorization to master them. This framework guides learners through a sequence of activities that involve listening, seeing, speaking, and writing. By engaging several senses at once, students build stronger mental connections and remember terms more accurately. Each step increases the level of participation required from the learner, allowing them to gain confidence while developing both receptive skills or understanding and productive skills or using the language.

Rather than presenting vocabulary as isolated items, the framework treats learning as an active process. Students encounter each term multiple times and in different ways, which helps prevent common problems such as incorrect pronunciation, confusion between similar terms, or forgetting spelling patterns. The approach is grounded in well-established learning theories that emphasize repetition, meaningful processing, and gradual skill development. Most importantly, it prepares students to use medical English in real academic and clinical situations, not just in written tests.

3.1. Step 1 “Recognize the Term”

In the first step, learners focus on recognizing the term through both sound and visual form. They hear the correct pronunciation while simultaneously seeing the word written in standard spelling and in phonetic symbols.

For example, the term *hypertension* is presented as audio together with its written form and its phonetic transcription:

- Audio: /ˌhaɪ.pəˈten.ʃən/
- Visual orthography: hypertension
- IPA: /ˌhaɪ.pəˈten.ʃən/

This combined exposure helps students connect how the word looks with how it sounds, which is especially important for medical terms whose pronunciation is not obvious from spelling alone.

This stage is mainly receptive, meaning that students are not required to produce the word yet. Instead, they build a clear mental model of the term. Seeing phonetic symbols also reduces confusion between letters and sounds, which is a common difficulty for learners whose first language uses different sound patterns. By establishing accurate recognition at the beginning, this step helps prevent the development of incorrect pronunciation habits that can be difficult to correct later.

Listening + Visual Recognition of Orthography + Phonetic Symbols

Table 1. Structure of Step 1 "Recognize the Term"

3.2. Step 2 "Produce the Sound"

In the second step, learners begin to actively produce the term. After hearing the pronunciation and seeing the written form, they repeat the word aloud while looking at it. This simple action significantly deepens learning because speaking requires coordination of listening, memory, and articulation. Students can also pay attention to stress patterns and rhythm, which are crucial for intelligible pronunciation in English.

Through repeated oral practice, learners start to feel more comfortable with complex multisyllabic terms. Words such as *myocardial infarction* become less intimidating once students practice saying them several times.

- Teacher plays audio: *myocardial infarction*
- Learners repeat: *myocardial infarction*
- Learners check stress pattern: /ˌmaɪ.oʊˈkɑːr.di.əl ɪnˈfɑːrk.ʃən/

This step strengthens the connection between sound and speech production and helps students internalize how the word should be pronounced in real communication. It also builds confidence, as students move from passive recognition to active participation.

Listening + Visual Recognition + Oral Repetition or Reading Aloud

Table 2. Structure of Step 2 "Produce the Sound"

3.3. Step 3 "Integrate the Form"

The third step introduces full multisensory practice by combining recognition with written production. Students listen to the term, read it, say it aloud, and then write it down. For example, when learning the word *bronchopneumonia*, they hear the pronunciation, see the spelling and phonetic form, repeat the word, and copy or write it.

- Term: *bronchopneumonia*
- Input: hear the audio + see word and IPA
- Output: repeat aloud + copy/write the word

This process engages both mental processing and physical movement, which strengthens memory.

Writing is particularly important for medical terminology because spelling accuracy matters in academic and clinical contexts. Many technical terms contain unfamiliar letter combinations that are difficult to remember without practice. By writing the word, students reinforce its visual structure and reduce the risk of future errors. At this stage, learners are actively integrating all language skills, which leads to deeper understanding and more stable retention.

Listening + Visual + Reading + Writing

Table 3. Structure of Step 3 "Integrate the Form"

3.4. Step 4 "Recall Independently"

The final step requires learners to produce the term independently based only on what they hear. The teacher or audio source says the word, and students write it without seeing the spelling or phonetic form. For instance, when they hear *electroencephalogram*, they must recall the correct spelling from memory.

- Teacher says: *electroencephalogram*
- Student writes: *electroencephalogram*

This activity tests whether the word has been fully learned rather than simply recognized.

Dictation is a powerful learning tool because it forces students to retrieve information from long-term memory. Successful completion of this step shows that learners can connect sound to spelling without support. It also mirrors real-life situations in which healthcare professionals must understand spoken terminology and record it accurately. By this stage, students demonstrate true mastery: they can recognize, pronounce, spell, and recall the term reliably.

Listening + Writing or Dictation and Recall

Table 4. Structure of Step 4 "Recall Independently"

3.5. Framework Summary

Together, the four steps form a coherent progression that guides learners from initial exposure to confident, independent use of medical terminology. The process begins with **Recognition**, in which students focus on accurate input by hearing the term, seeing its written form, and noticing its pronunciation features. This stage builds a reliable foundation and reduces the risk of misunderstanding or incorrect learning from the start. Once learners can clearly identify the term, they move to **Repetition**, where they practice saying it aloud while continuing to receive auditory and visual support. This stage strengthens pronunciation, rhythm, and stress patterns, helping students become more comfortable producing complex multisyllabic words.

The third stage, **Integration**, expands practice across multiple language skills. Students listen to the term, read it, pronounce it, and write it, combining receptive and productive processes in a single activity. This multi channel engagement deepens processing and supports both spelling accuracy and pronunciation control. Finally, the process culminates in **Production**, where learners reproduce the term independently from memory, typically through dictation or recall tasks. At this point, students demonstrate true mastery because they can recognize, pronounce, spell, and retrieve the word without external support.

Breaking learning into these manageable stages makes complex medical vocabulary less intimidating and more achievable. Each step reinforces the previous one, creating a cycle of repetition, consolidation, and increasing independence. Through this spiral progression of Recognition, Repetition, Integration, and Production, vocabulary learning shifts from passive exposure to active engagement, enabling students to retain terms longer and use them with greater confidence in academic and clinical communication.

4. Discussion

4.1 Multisensory Learning and Vocabulary Retention

One of the most compelling reasons to adopt a multisensory framework for medical vocabulary learning lies in its potential to enhance **long-term retention**. According to Paivio's (1986) *Dual Coding Theory*, information processed through both verbal and non-verbal systems is more effectively stored in long-term memory. In the case of medical terminology, auditory input (hearing pronunciation) and visual input (orthography and IPA transcription) create parallel cognitive representations, reinforcing recall. Learners who encounter the term *electroencephalogram* through listening, seeing, and writing are not merely memorizing a string of letters; they are building layered associations that strengthen mental connections.

Empirical evidence also supports this claim. Shams and Seitz (2008) reviewed research across disciplines and concluded that multisensory training enhances both learning speed and memory durability. Applied to the present framework, Step 1 and Step 2 maximize this effect by presenting terms through simultaneous auditory and visual

channels, while Step 3 and Step 4 extend learning through active kinesthetic reinforcement. Such a design aligns with Nation's (2001) principle that vocabulary acquisition requires multiple encounters in varied contexts to transition from short-term recognition to durable mastery.

4.2 Pronunciation Accuracy and Orthographic Awareness

Pronunciation and spelling remain persistent challenges in ESP contexts. Kelly (2000) warned that learners often fossilize incorrect pronunciations when left without explicit phonetic guidance. The integration of **phonetic transcription (IPA)** at Step 1 of the framework directly addresses this issue by linking orthographic form to standardized phonetic symbols. Learners are less likely to mispronounce *gastroenterology* when they are explicitly shown its IPA form /ˌgæs.trəʊˌen.təˈrɒl.ə.dʒi/.

Step 2 builds on this by requiring learners to repeat terms aloud, activating both auditory and articulatory memory. Craik and Lockhart's (1972) *Levels of Processing* framework emphasizes that deeper cognitive processing, such as articulation, enhances retention. This explains why oral repetition, although sometimes dismissed as "mechanical," is cognitively powerful. By Step 3, the integration of writing consolidates orthographic precision, ensuring learners can spell terms such as *anaphylaxis* correctly—a skill essential in clinical and academic communication.

The importance of linking pronunciation and orthography is further underscored by Bosher and Smalkoski (2002), who found that nursing students' difficulties with medical English often stemmed from mispronunciation and misspelling of technical vocabulary. These issues hindered both their academic performance and their ability to communicate in clinical settings. The multisensory framework therefore directly addresses a barrier that has been empirically identified in ESP learning contexts.

4.3 Depth of Processing and Productive Vocabulary Knowledge

Vocabulary knowledge cannot be reduced to recognition alone. Nation (2001) makes a clear distinction between **receptive knowledge** (understanding when encountered) and **productive knowledge** (ability to actively use). Step 4 of the framework—listening and writing through dictation—ensures that learners achieve productive mastery by recalling terms without visual prompts. This stage is supported by retrieval practice research, which highlights that recall tasks, compared to recognition tasks, produce stronger and more durable memory traces (Roediger & Butler, 2011).

Craik and Lockhart's (1972) theory further explains why this progression is effective: shallow processing (Step 1 recognition) is necessary but insufficient for long-term retention. By Step 2 and Step 3, the act of reading aloud and writing increases the depth of processing, while Step 4 requires recall, the deepest form of engagement. Each stage in the framework is therefore not redundant but cumulative, scaffolding learners toward independent and confident use of medical terminology.

4.4 Relevance to ESP and Medical English Contexts

The specific context of medical English magnifies the importance of accurate vocabulary acquisition. Dudley-Evans and St John (1998) argued that specialized lexis is central—not peripheral—to ESP courses, as it underpins learners' ability to function in professional domains. Yet, as Hutchinson and Waters (1987) pointed out, ESP instruction often relies on adaptation of general English methods without sufficient tailoring to specialized needs. The four-step multisensory framework fills this gap by focusing explicitly on the challenges posed by medical terms: multisyllabic length, complex stress patterns, and morphology derived from Latin and Greek.

Bosher and Smalkoski's (2002) study of immigrant nursing students in the United States demonstrated how inadequate mastery of medical vocabulary hindered their academic progression and professional communication. This provides strong justification for a model that not only introduces terms but ensures their mastery through multiple sensory and cognitive pathways. For example, when a Vietnamese student learns *myocardial infarction* by hearing, repeating, writing, and recalling, they are preparing not only for classroom success but also for clinical accuracy.

4.5 Digital Applications and Multimedia Learning

The framework also aligns with research on multimedia learning. Mayer (2001) argued that learners benefit most from multimodal input when words and visuals are presented together in a manner that reduces extraneous load. The four-step framework can be digitized into **Learning Management Systems (LMS)** such as Moodle, where audio, orthography, and IPA can be embedded for Step 1, recording tools for Step 2, writing tasks for Step 3, and automated dictation tests for Step 4.

Mobile applications further extend these possibilities. Liakin, Cardoso, and Liakina (2015) found that mobile speech recognition technology significantly improved learners' oral accuracy. Such tools can support Step 2 by allowing learners to receive immediate feedback on pronunciation. Similarly, spaced repetition systems like Anki can be adapted for multisensory practice by combining sound files, text, and writing prompts. These technologies transform the framework from a classroom-only method into a scalable approach for blended and online learning environments.

Looking forward, innovations in **Augmented Reality (AR)** and **Virtual Reality (VR)** could further enhance the multisensory approach. AR flashcards could simultaneously display a medical term, play its pronunciation, and provide writing prompts. VR hospital simulations could immerse learners in clinical contexts where they must listen to, repeat, and write medical terms in real time. Such applications align with Shams and Seitz's (2008) call for integrating multisensory methods into educational technology, ensuring engagement and immersion.

4.6 Challenges and Limitations

Despite these benefits, challenges remain. Time constraints are significant in medical curricula, which are often overloaded with technical content. Integrating a four-step cycle for every term may be impractical unless carefully streamlined. Teachers must decide which high-frequency or high-risk terms warrant multisensory treatment, following Nation's (2001) principle of prioritizing vocabulary by frequency and usefulness.

Teacher expertise is another limitation. Oxford (2001) emphasized that learners possess diverse styles and strategies, and instructors must be trained to adapt teaching accordingly. Many ESP instructors lack confidence in teaching pronunciation or using IPA (Kelly, 2000). Without institutional support for professional development, the consistency and accuracy of implementation may vary.

Technological disparities also pose challenges. While some institutions have robust LMS and access to mobile technology, others may face infrastructural barriers. Low-tech adaptations—such as teacher-led dictation and printed IPA charts—must be considered to ensure equity. Furthermore, accessibility for students with hearing or visual impairments requires thoughtful adjustments to maintain inclusivity.

Finally, learner perception may initially resist repetition or dictation. Dudley-Evans and St John (1998) note that ESP learners often prefer content-focused tasks to language drills. Teachers must therefore frame multisensory practice as a professional skill-building exercise rather than mechanical repetition.

4.7 Implications for Internationalization and Global Practice

The implications of this framework extend beyond individual classrooms. As universities worldwide adopt **English-Medium Instruction (EMI)** and seek to internationalize, mastery of specialized terminology becomes critical for global mobility. Accurate pronunciation and spelling of medical terms enable students to read international journals, participate in conferences, and collaborate with peers across borders.

In contexts such as Vietnam, China, or Latin America, where English is taught as a foreign language, multisensory strategies provide a practical pathway to bridging the gap between local education and global professional standards. The framework supports not only individual learner success but also institutional goals of producing graduates who are internationally competent. By systematizing vocabulary learning, it strengthens the foundation for broader educational transformations in medical and health sciences.

4.8. Summary of Discussion

The discussion has demonstrated that the four-step multisensory framework offers substantial pedagogical, technological, and institutional benefits. Supported by established theories such as Paivio's *Dual Coding Theory*, Mayer's *Multimedia Learning*, and Craik and Lockhart's *Levels of Processing*, as well as empirical findings by Nation (2001), Bosher and Smalkoski (2002), and Liakin et al. (2015), the model is grounded in evidence. It ensures vocabulary retention, pronunciation accuracy, spelling mastery, and learner confidence. It can be adapted for classroom, digital, and immersive learning environments, and it aligns with internationalization agendas in higher education. While challenges exist in time, training, and technology, these are outweighed by the framework's capacity to transform medical vocabulary learning from rote memorization into active, multisensory mastery.

5. Conclusion

The acquisition of medical vocabulary represents one of the most formidable challenges for learners of English for Specific Purposes (ESP). Terms of Latin and Greek origin are often long, morphologically complex, and phonetically difficult, demanding a pedagogical approach that goes beyond rote memorization. This paper has

proposed and discussed a four-step multisensory framework "listening and visual recognition with phonetic transcription, oral repetition, integration of reading and writing, and dictation-based recall" as a systematic method for enhancing medical terminology learning.

The framework is firmly grounded in cognitive and educational theory. Paivio's (1986) *Dual Coding Theory* and Mayer's (2001) *Cognitive Theory of Multimedia Learning* provide strong justification for combining auditory and visual channels. Craik and Lockhart's (1972) *Levels of Processing* model explains the effectiveness of moving from shallow recognition to deeper recall, while Nation (2001) emphasizes the importance of achieving both receptive and productive knowledge. Evidence from ESP research, including Bosher and Smalkoski (2002) and Dudley-Evans and St John (1998), highlights the practical consequences of inadequate vocabulary mastery in medical education, reinforcing the need for innovative approaches such as this framework.

The discussion has demonstrated that multisensory integration provides not only improved retention, pronunciation, and spelling accuracy, but also enhanced learner confidence and engagement. Furthermore, the framework is adaptable for both traditional classrooms and digital learning platforms, including LMS, mobile applications, and potentially AR/VR environments. Although challenges remain—particularly in time allocation, teacher training, and technology access—the potential benefits outweigh the limitations.

Looking forward, future research should empirically test the framework through classroom interventions and controlled studies, measuring outcomes such as recall rates, pronunciation accuracy, and learner attitudes compared with traditional methods. Such research would provide valuable evidence to refine and validate the model. Ultimately, the framework offers a scalable, evidence-based approach that supports learners not only in academic settings but also in preparing for globalized healthcare communication. In doing so, it aligns with the broader goals of sustainability, digital transformation, and internationalization in higher education.

6. References

Bosher, S., & Smalkoski, K. (2002). From needs analysis to curriculum development: Designing a course in health-care communication for immigrant students in the USA. *English for Specific Purposes, 21*(1), 59–79. https://doi.org/10.1016/S0889-4906(01)00002-3

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for Specific Purposes: A multi-disciplinary approach*. Cambridge University Press.

Kelly, G. (2000). *How to Teach Pronunciation*. Longman.

Liakin, D., Cardoso, W., & Liakina, N. (2015). Mobile-assisted language learning: Affordances and limitations of speech recognition technology. *ReCALL, 27*(3), 298–318. https://doi.org/10.1017/S0958344015000077

Mayer, R. E. (2001). *Multimedia Learning*. Cambridge University Press.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge University Press. https://doi.org/10.1017/CBO9781139524759

Oxford, R. (2001). *Language Learning Styles and Strategies*. Mouton de Gruyter.

Paivio, A. (1986). *Mental Representations: A Dual Coding Approach*. Oxford University Press.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20–27. https://doi.org/10.1016/j.tics.2010.09.003

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in Cognitive Sciences, 12*(11), 411–417. https://doi.org/10.1016/j.tics.2008.07.006