

Ivana M. Krsmanović
University of Kragujevac
Faculty of Technical Sciences Čačak
Department of Social Sciences and Humanities
ORCID 0000-0003-4793-628X

Marija D. Nikolić
University of Kragujevac
Faculty of Technical Sciences Čačak
Department of Information Technologies
ORCID 0000-0002-1049-4708

USING MULTIMODAL DIGITAL TOOLS IN ESP CLASSROOMS: PEDAGOGICAL APPLICATIONS OF VOICETHRED

This paper explores the pedagogical potential of *VoiceThread*, a multimodal digital tool, within the framework of computer-assisted language learning (CALL) in English for Specific Purposes (ESP) instruction. Responding to the growing demand for professionally oriented communicative competence, the study adopts a practice-based, descriptive-analytical approach. It demonstrates how *VoiceThread* can be systematically integrated into ESP classrooms to support discipline-specific speaking tasks through asynchronous, multimodal interaction. Drawing on sustained classroom observation and instructional practice, the paper illustrates task design principles and the role of teacher mediation in CALL-supported ESP contexts. Observational insights point to several pedagogical benefits, including increased learner engagement, greater willingness to participate in spoken tasks, enhanced awareness of professional discourse conventions, and opportunities for reflection and peer interaction beyond classroom constraints. The paper argues that multimodal CALL tools such as *VoiceThread* can function as effective pedagogical environments for ESP instruction, offering practical guidance for instructors seeking to enrich spoken communication and professional language development.

Keywords: English for Specific Purposes (ESP), VoiceThread, Multimodality, Computer-assisted Language Learning (CALL), Digital Pedagogy.

1. INTRODUCTION

English for Specific Purposes (ESP) has long been concerned with aligning language instruction to the communicative demands of particular disciplines and professional communities. In contemporary contexts, however, these demands are no longer realized primarily through linear,

print-based texts. Instead, professional communication increasingly unfolds through multimodal approaches that integrate language with visual, spatial, audio, and digital resources, ranging from technical diagrams and data visualizations to instructional videos, simulations, and collaborative platforms. As a result, ESP pedagogy must move beyond a narrow focus on linguistic competence to address how meaning is constructed, negotiated, and performed across multiple semiotic modes.

Within this landscape, multimodality offers a powerful theoretical and pedagogical lens for rethinking ESP. Drawing on social semiotic perspectives, multimodality foregrounds the purposeful selection and interpretation of semiotic resources in relation to specific communicative goals, audiences, and contexts. For ESP learners, this entails not only understanding specialized lexis and genres, but also developing the capacity to interpret and produce multimodal texts that reflect contemporary professional practices. Technical reports, engineering briefings, innovation proposals presentations, and business pitches, for example, rely on the interplay of verbal explanation, visual representation, and embodied action.

At the same time, the digital transformation of higher education has expanded the spaces in which ESP learning takes place. Learning environments now encompass asynchronous discussion tools, collaborative platforms, and media-rich tasks that blur the boundaries between academic, professional, and everyday communication. In this setting, multimodal ESP pedagogies position learners as active designers of meaning, drawing on their disciplinary knowledge, technological repertoires, and lived experiences to engage with authentic communicative challenges. This paper argues that integrating multimodality into ESP is no longer optional but essential for preparing learners to participate effectively in contemporary professional communities, and it explores how multimodal design principles can be operationalized within ESP curricula in higher education.

Despite a growing body of research documenting the pedagogical affordances of *VoiceThread* in English as a Foreign Language (EFL) contexts, its uptake within English for Specific Purposes (ESP) remains limited. This field deserves more scholarly attention, given that ESP learners are precisely those who must navigate complex multimodal genres and professional communication practices. The relative absence of *VoiceThread* usage in ESP pedagogy is therefore less a matter of suitability than of unrealized potential. Addressing this gap, the purpose of this paper is to demonstrate how *VoiceThread* can be systematically integrated into ESP instruction and to articulate why its multimodal, asynchronous affordances make it especially well-suited to contemporary ESP learning and teaching contexts.

2. MULTIMODALITY AND COMPUTER-ASSISTED LANGUAGE LEARNING

In the digital age, perhaps more than at any other moment in human history, communication has expanded beyond spoken and written language to encompass rich constellations of visual, spatial, gestural, and digital modes that operate together to produce meaning.

Multimodality, both in communication and education, refers to the use of more than one mode of communication in a text to create meaning. It examines how diverse semiotic resources, such as visual, linguistic, spatial, and gestural modes, are organised to make meaning (Lim et al., 2022), and describes how each mode contributes to the overall meaning of a multimodal text (Donaghy 2024). Closely related to it is the concept of multimodal pedagogy, which denotes the intentional “design of learning experiences through the selection, sequencing, and integration of multiple modes to represent curriculum content” (Bezemer, Kress 2015).

As a concept, multimodality gained significant popularity among scholars between 2019 and 2021 (Hendrawaty et al. 2024). Only recently has Multimodality and English Language Learning (MELL) been defined as an interdisciplinary, emerging field at the intersection of multimodality and language learning in which English, as a global lingua franca, mediates a wide range of social interactions with texts, such as gaming and social media practices (Diamantopoulou, Ørevik 2025). Broadly speaking, Multimodal English Language Learning (MELL) is closely related to Computer-Assisted Language Learning (CALL) as an applied linguistic field, in that both employ digital technologies to support language learning. CALL has long been perceived as an efficient supporter of the EFL and ESP teaching-learning process, significantly adding to students’ overall improvement in English language competence, while at the same time empowering them to engage more in self-study (Krsmanović 2021).

Current trends, benefits and challenges of a multimodal approach in ELT are nicely outlined in a systematic literature review that analyses 27 scholarly publications (Hendrawaty et al. 2024). The review confirms that multimodal approach to ELT enhances not only language-related skills, such as academic writing skills and vocabulary mastery, but also improves 21st century skills such as develop creativity, critical thinking, dimensions of literacy. The study findings further indicate that multimodality in ELT is superior to monomodal methods. For example, of the articles that measured efficacy, six out of seven found that the multimodal activities were more effective than traditional monomodal (single-mode) strategies. Furthermore, multimodal approach to teaching English was found to significantly boost student motivation and engagement, while fostering positive attitudes toward learning, promoting learner autonomy and inclusion, and facilitating various learning styles by providing diverse pathways to meaning. What might be potential limitations of the approach, but com-

mon to any other CALL tool or app, is a need for curriculum redesign to keep pace with advanced technologies, and additional teacher training, so that language educators can improve their multimodal literacy and skills in selecting culturally relevant materials (Hendrawaty et al. 2024).

Kieran Donaghy defines the Multimodal Approach as an innovative pedagogical framework designed to reflect the realities of communication in our digitally and visually mediated world (Donaghy 2024). Alongside traditional reading, writing, listening, and speaking, multimodal approach to teaching English integrates two essential 21st-century skills; viewing (an active, interpretive process where learners decode various modes while responding emotionally and critically to a text), and representing (the ability to communicate ideas by creating multimodal texts, such as infographics, video essays, or memes) (Donaghy 2024).

Donaghy argues that the key educational benefits of a multimodal approach include deeper learning, enhanced empathy and human connection, greater inclusion and accessibility, and improved vocabulary retention. Multimodal instruction supports deep learning by engaging both verbal and visual processing channels, enabling learners to retain information more effectively. At the same time, multimodal texts incorporate affective cues that help learners interpret emotional nuance, fostering empathy and intercultural awareness. From an equity perspective, multimodality offers multiple pathways to meaning, increasing accessibility for learners with reading difficulties or sensory impairments by enabling engagement through visual and auditory modes rather than print alone. Finally, vocabulary learning is strengthened through the systematic recycling of key terms across modes and tasks, ensuring repeated and meaningful exposure that promotes long-term retention (Donaghy 2024).

3. INTRODUCING THE MULTIMODAL TOOL VOICETHREAD

VoiceThread.com, founded in 2005, is a US educational technology company that provides a cloud-based multimedia platform for collaborative communication. *VoiceThread* is a user-friendly multimodal ICT tool which allows users to upload images, videos, documents, or presentations and comment on them through recorded voice, webcam, or text, by using what is called a “thread”. A “thread” can be shared within a closed group of learners as well as with other educators in a public mode. The key feature of the tool, and its major multimodal quality, is the possibility to engage with a text/presentation/photo/video by providing comments on them through recorded voice, webcam, or text. Such collaborative learning experiences are then saved via these “threads” and can be accessed any time, aligning with asynchronous and blended learning models, across disciplines and devices. The system integrates with major learning management systems (LMSs) such as Canvas and Moodle, and offers 99.9% uptime for institu-

tional clients. Its *VoiceThread* Universal interface ensures accessibility for screen-reader users and compliance with Section 508 standards.

Initially conceived as a consumer media-sharing tool, *VoiceThread* gained early adoption among teachers who used it to amplify student voices in digital classrooms. It is now widely used in K–12 and higher education for presentations, reflections, and peer feedback. Educators cite its ability to promote social presence, empathy, and engagement in online environments where text-only interactions can feel impersonal (Acosta, Parra 2023).

VoiceThread emphasizes inclusive design, providing automatic captioning through Amazon Web Services technologies such as Amazon Transcribe and text-to-speech features via Amazon Polly. Its “Universal” mode enables full participation by users with visual impairments, while the mobile app extends collaboration across platforms and bandwidth conditions. Finally, *VoiceThread* has been widely recognized as a versatile tool for asynchronous communication that blends human voice, visuals, and interactivity to enrich online learning experiences.

3.1. PREVIOUS RESEARCH

A growing body of research has examined the pedagogical use of *VoiceThread* in English language teaching, with much of this work situated within general ELT and EFL contexts. More limited, but emerging, research has begun to explore its application in English for Specific Purposes (ESP).

For example, Bodis et al. (2020) identify several pedagogical benefits of *VoiceThread* used in ESP instruction, such as enhanced learner autonomy, increased feedback literacy, increased engagement and motivation, development of learning community, as well as strengthening the information and communication technology (ICT) skills for students. This is in line with studies conducted by Utami and Santosa (2023), and Zemlyanova et al. (2021), who also reported comparable benefits, only in EFL contexts.

Among many other pedagogical affordances, *VoiceThread* can be easily embedded in different LMS, which allows students to learn and share within a safe learning environment, similar to face-to-face, which was confirmed in several studies (Bodis et al. 2020; Ward et al. 2017). It was also found that *VoiceThread* can reduce student isolation, sharpen communication skills, and increase emotional connection when students receive audio feedback from instructors. It also allows for asynchronous reflection, giving students more time to think before they respond compared to face-to-face discussions (Chen L. 2025; Ward et al. 2017).

Chen M. (2025) found *VoiceThread* to be beneficial in a flipped second-language classroom context. The tool was shown to function as an effective multimedia platform for asynchronous interaction and the delivery of lecture content, supporting self-paced learning, the development of a growth mindset, and knowledge internationalization. In addition to these

affordances, teachers reported that *VoiceThread* enabled them to review and update their own knowledge of complex topics while simultaneously enhancing instructional innovation, leading to more student-centered and reflective course planning. Similarly, Utami et al. (2023) found that the use of *VoiceThread* has a consistently positive effect on the development of speaking skills among English as a Foreign Language (EFL) learners. Their study reported significant gains in oral presentation skills, fluency, and accuracy, and further demonstrated that *VoiceThread* is effective for IELTS preparation and for fostering critical thinking alongside language proficiency.

Zemlyanova et al. (2021) conducted a study with a sample of 44 Russian university students in non-linguistic departments. The study found that using mobile *VoiceThread* had a statistically significant positive effect on the development of oral English skills compared to traditional methods. Improvements were noted across five analytic criteria: fluency, content, pronunciation, grammar, and vocabulary. The study emphasized that the multimodal nature of the tool provided students with numerous benefits, one of them being related to its asynchronous nature which encouraged more intensive study habits. Namely, students were engaged in additional oral practice because they were allowed to record and re-record their responses as many times as they liked before final submission. This also led to a reduced speaking anxiety as shy students were allowed to edit their speech and avoid the embarrassment of making public mistakes in real-time, which provided additional safety for shy learners.

Finally, when talking about CALL, prior research in online and hybrid teaching and learning has demonstrated that learning experiences, learning outcomes and learner satisfaction are significantly increased if the interaction is intentionally designed through activation and sustain of three presences, namely Social Presence, Teaching Presence and Cognitive Presence, defined under the framework of the Community of Inquiry (CoI) by Garrison, Anderson and Archer in 1999 (Krsmanović et al. 2025). Compliant to that, research studies confirm that *VoiceThread* has been shown to foster participation practices that further strengthen the three CoI, which makes it a resourceful tool for successful ELT (deNoyelles et al. 2014).

Similarly, Chen L. (2025) investigated the role of *VoiceThread* grounded in CoI framework and English as a Medium of Instruction (EMI), proposing a mentoring approach to the tool use inside and outside the classroom. The study found that *VoiceThread* is valued for its ability to transmit Social Presence through tone of voice, cadence, and body language (elements that account for approximately 93% of communication and are often lost in text-only online courses). What is more, multimodality was found to be beneficial, as features such as visual images, sound effects, and video overlays are highlighted as effective tools for Generation Z students, who prefer video-based learning over textbooks (Chen L. 2025). Finally, combining computer-mediated techniques like *VoiceThread*

with EMI demonstrated to offer a more authentic and participatory way to improve student retention (Chen L. 2025).

4. DEMONSTRATING THE USE OF VOICETHREAD IN ESP INSTRUCTION

To illustrate the pedagogical use of *VoiceThread* in English for Specific Purposes (ESP) instruction, this section presents a representative activity drawn from a *Technical English* course designed for master's-level students of Mechanical and Electrical Engineering at the Faculty of Technical Sciences Čačak. The activity exemplifies how a multimodal, asynchronous environment can be employed to support discipline-specific oral communication aligned with professional engineering practices. The description is based on classroom practice and observational insights.

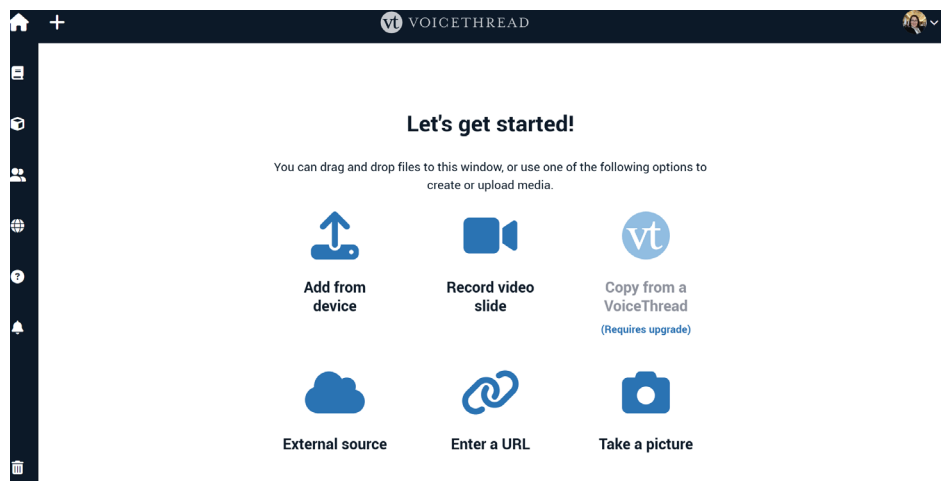


Figure 1 *VoiceThread* user layout

First, the application enables the instructor to upload or create a range of media, including photographs, PowerPoint presentations, videos, or external links (Figure 1). Once the media content is uploaded, a “thread” is created, which can be titled, organized, and shared with students. This “thread” serves as the central space for multimodal interaction and task completion and can be additionally altered and operationalized.

For further adaptation, the instructor can access the *Settings* panel to configure participation and interaction options (Figure 2). These include controlling who is permitted to respond to the “thread”, selecting the permitted response modes (voice, video, or text), and determining whether comments are public, moderated, or restricted. Such settings allow instructors to align the activity with specific pedagogical objectives,

for example by encouraging oral production, managing peer feedback, or scaffolding participation in ESP speaking tasks.

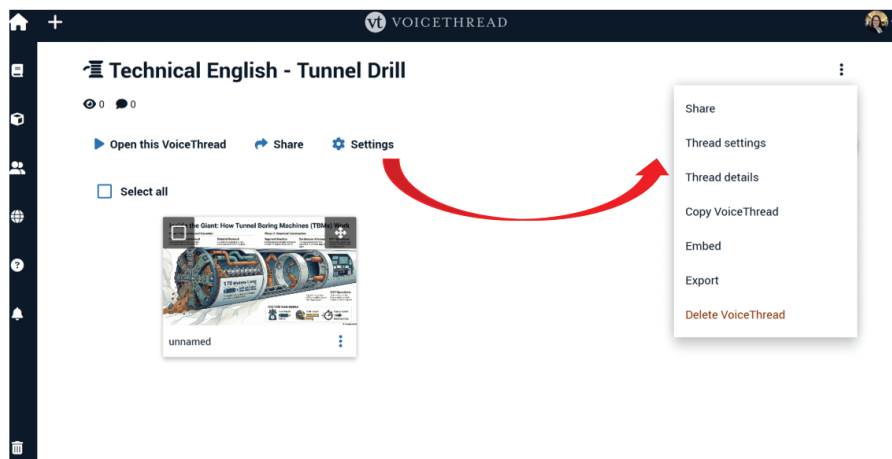


Figure 2 Uploading a new thread and user options

All “threads” created by the instructor are easily visible and accessible from the Home page, allowing for efficient management of course materials. Importantly, once created, “threads” can be reused, modified, and adapted for different instructional purposes or cohorts, enabling instructors to refine tasks over time and maintain continuity across ESP courses (Figure 3).

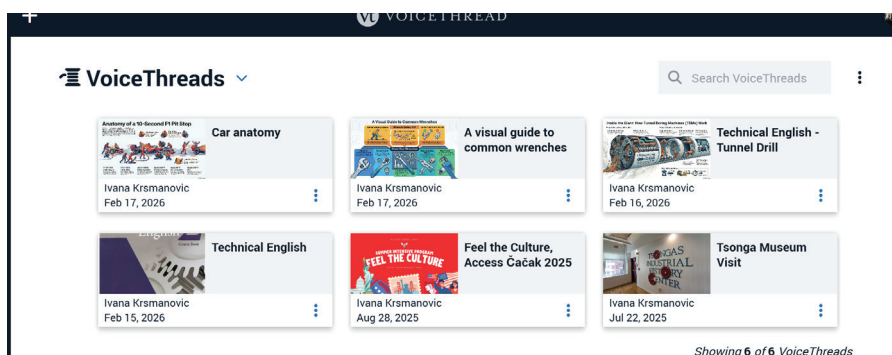


Figure 3 Homepage layout

When a “thread” is opened, the application offers four modes of engagement with the material: written text, voice messages, video messages, and the upload of additional resources (Figure 4). These multiple response modes enable learners to select the form of interaction that best supports their communicative intent and technical explanation. The instructor can use all the modes to provide the lesson instructions, lead the class or add additional resources.

For illustrative purposes, the activity presented here is entitled *Inside the Giant: How Tunnel Boring Machines (TBMs) Work*, and as a key prompt it uses an infographic created with NotebookLM (Studio option: *Infographic*). The infographic was generated on the basis of a YouTube video entitled *How Tunnel Boring Machines Work*, which introduces a complex engineering process of a tunnel boring, through visual and verbal explanation (Figure 5). This multimodal prompt was deliberately selected to support ESP learners in producing structured, discipline-specific oral commentary, combining visual input with technical vocabulary and explanatory discourse.

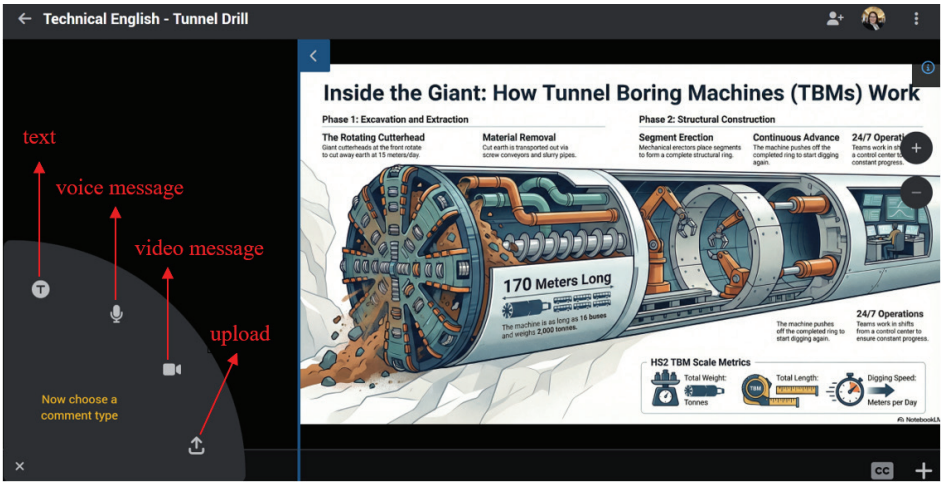


Figure 4 Tunnel Drill “thread” with four response options: text, voice, video and upload

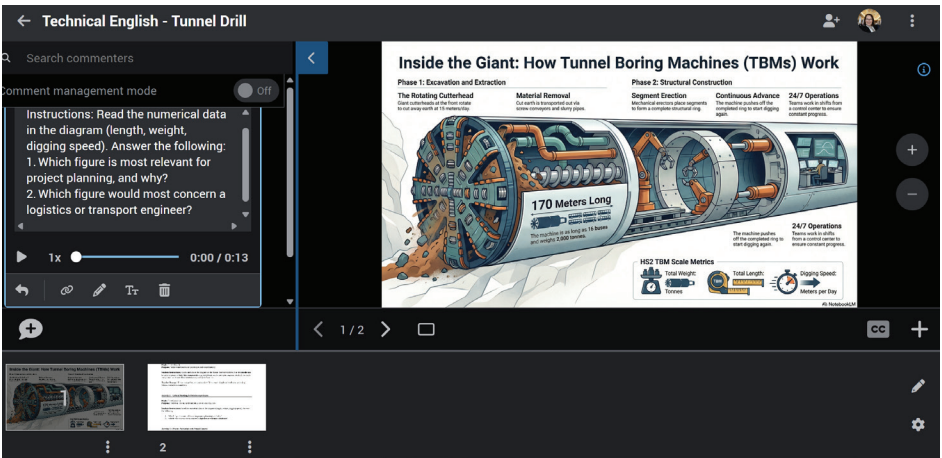


Figure 5 Layout of a posted text response

A “thread” can consist of different types of learning materials, and annotation and commenting functions can be enabled or disabled individually for each item. This flexibility allows the instructor to regulate

interaction, guide learner focus, and tailor participation according to the pedagogical aims of a given ESP task. The layout of a video instruction given by the instructor is given in the left-hand upper corner of the “thread” (Figure 6).

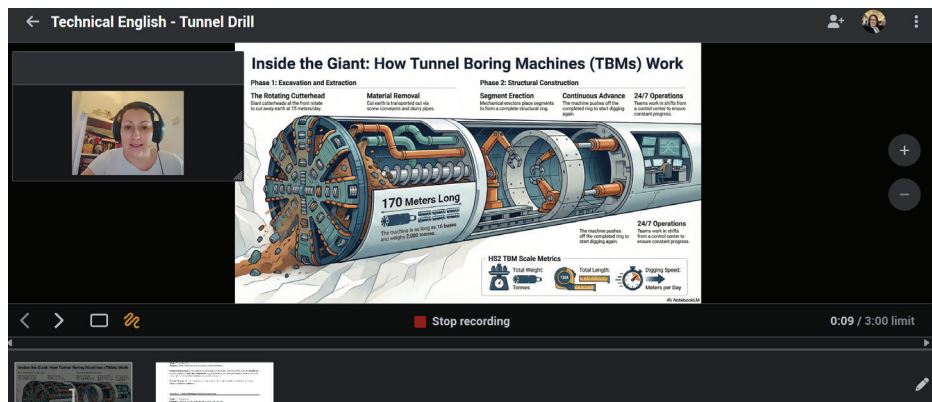


Figure 6 Video response within a thread

For written (text) comments, the tool includes an automatic (embedded) reading aid that uses dynamic text tracking, whereby the text is visually highlighted in real time as it is read (Figure 7). This feature supports reading fluency and comprehension by guiding learners' attention through the text and reinforcing the connection between written form and meaning, which can be particularly beneficial in ESP contexts where dense technical language is involved. It should be noted that the examples presented in this section illustrate instructor-designed activities only; student contributions are not included.

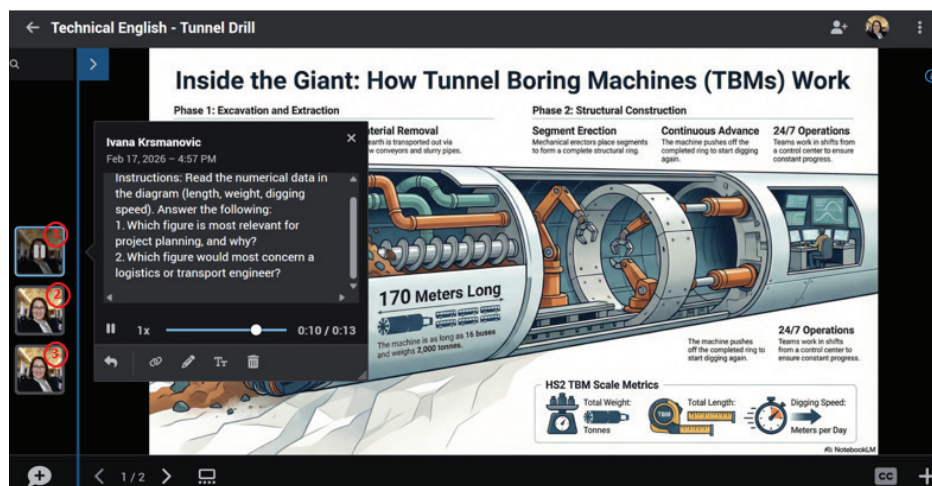


Figure 7 Dynamic text tracking layout with positioning of all responses (1, 2, 3) at a left-hand side, lined vertically

5. INSTRUCTIONAL DESIGN PRINCIPLES AND LESSONS LEARNED

The instructional design of the *VoiceThread* activity *How Tunnel Boring Machines Work* presented here was guided by several core principles aligned with ESP pedagogy and multimodal learning:

a) Integration of Multiple Semiotic Modes: As it has been pointed out earlier, effective task design must move beyond linguistic competence to address how meaning is constructed across multiple modes. Instructors should purposefully select and sequence resources in a “thread”, such as technical diagrams, videos, and verbal explanations, to reflect authentic professional communication. In our case, the topic was introduced through a YouTube video, which served as the initial input. This was followed by the generation of an infographic to visualise key concepts, after which a sequence of structured speaking activities was implemented to support students’ oral production and discipline-specific communication (a Word or PDF handout).

b) Scaffolding through teacher mediation: The instructor plays a critical role in configuring the learning environment to guide student focus. By utilizing the tool’s settings, such as moderating comments or selecting specific response modes (voice, video, or text), instructors can provide the necessary pedagogical scaffolding to help learners navigate complex technical tasks. For this class activity, the instructor provided scaffolding by using all four engagement modes, text, voice, video, and resource uploads, to provide lesson instructions, lead the class discussion, or add supplementary resources that support the learners’ understanding.

c) Fostering a Community of Inquiry (CoI): To activate Social, Teaching, and Cognitive Presences, interaction must be intentionally designed. The design should prioritize asynchronous “threads” that allow for reflection, peer interaction, and the transmission of Social Presence through voice and body language, which are often lost in text-only environments. For these reasons, particular emphasis was placed on the transmission of Social Presence through vocal cadence, intonation, facial expression, and body language, communicative cues that are essential in professional discourse yet often absent from text-only environments. To support these aims, we deliberately prioritised audio and video responses over written comments, encouraging students to engage in spoken, discipline-specific communication that more closely mirrors authentic professional interaction.

d) Learner-centred design for autonomy: Tasks should position learners as “active designers of meaning”, allowing them to choose the response modes that best support their communicative

intent. Providing opportunities to record and re-record responses promotes intensive study habits and reduces speaking anxiety. To do so, our students were given flexibility in choosing between audio and video responses and were encouraged to rehearse, review, and refine their contributions before posting, thereby fostering autonomy, reflective practice, and greater control over their spoken professional discourse.

5.1. OBSERVATIONAL INSIGHTS AND PEDAGOGICAL BENEFITS

Based on sustained classroom observation and instructional practice presented above, several pedagogically meaningful insights emerged from the integration of *VoiceThread* into ESP instruction. One of the most prominent observations was an increase in learners' willingness to engage in spoken production. The asynchronous and multimodal nature of the platform appeared to reduce performance pressure, allowing students to plan, rehearse, and refine their responses before sharing them. This is particularly beneficial in ESP contexts, where learners are often hesitant to speak due to the perceived complexity of technical content and terminology. Moreover, *VoiceThread* facilitated rehearsed professional speaking by allowing students to record and re-record their responses multiple times prior to final submission. This affordance is especially valuable for ESP learners who must master precise technical lexis, as it enables less confident students to edit their output and avoid the anxiety they often feel when faced with public speaking.

Another observation we can make here is that students demonstrated heightened awareness of discipline-specific discourse conventions. The requirement to explain technical concepts via recorded voice or video messages, supported by visual prompts given via *VoiceThread*, encouraged more structured, audience-oriented communication. This process was further reinforced through repeated exposure to discipline-specific discourse across multiple modes. Namely, from what we have observed, multimodality enabled the systematic recycling of key terms through spoken explanation, visual representation, and written input, strengthening lexical consolidation. Moreover, what we observed is that master's students showed a clear preference for precise infographics and similar graphical data used as a prompt in *VoiceThread*, as they associate such visual data as a distinctly engineering-oriented mode of communication.

Thirdly, peer interaction also played a significant role in the *Tunnel Drill* activity. The threaded, multimodal structure of the platform created a collaborative learning environment in which students engaged with one another's contributions through voice and video comments. At the same time, instructor audio feedback strengthened Teaching Presence and guided learners' attention toward completion of the given tasks. This extended engagement with spoken discourse supported deeper cognitive processing

and contributed to the development of professional communicative competence in ESP settings.

Finally, as an asynchronous activity conducted within the course, the “thread” could be revisited at any stage, allowing students to review previous contributions, revise their responses, and engage in ongoing reflection, thereby supporting continuous language improvement. This was very useful both for students who attended the class and who relied on the “thread” afterwards to compensate for missed sessions.

6. CONCLUSION

As digital transformation continues to reshape higher education, especially with the advent of AI, adopting multimodal CALL environments has become a fundamental requirement for equipping ESP learners with the communicative competence required for their future careers. The practical application of the *VoiceThread* tool within a Technical English course for engineering students illustrates how intentional instructional design and teacher mediation can guide learners through complex discipline-specific tasks, fostering rehearsal, reflection, and peer-supported learning. By integrating visual, auditory, and textual modes, the platform enables learners to engage deeply with professional discourse and practice the ESP terminology. Such multimodal interaction not only enhances learner autonomy and confidence but also mirrors authentic professional environments, highlighting the transformative potential of CALL in preparing ESP students for the demands of their respective fields.

Classroom observations from the teaching activity Tunnel Drill suggest that *VoiceThread* offers a pedagogically rich space in which ESP learners can engage in rehearsed professional speaking, reduce performance-related anxiety, and develop greater awareness of disciplinary discourse conventions. The combination of visual prompts, audio and video responses, and peer interaction can facilitate deeper engagement with technical content and promote reflective learning beyond the temporal constraints of the classroom. Importantly, the tool’s affordances can easily enable instructors to maintain Teaching presence through task design, scaffolding, and targeted feedback, while also fostering social and cognitive presence among learners.

While the paper does not claim empirical validation of learning outcomes, it argues that pedagogically informed demonstrations such as this provide valuable insight into how multimodal tools can function as instructional environments rather than mere technological add-ons. The findings underscore the relevance of multimodal Computer-assisted Language Learning for ESP contexts and highlight the need for further research that empirically examines the long-term impact of such tools on professional communicative competence across disciplines.

REFERENCES

- Acosta, Parra 2023: K. Acosta, & E. H. Parra, *Fostering Engagement with VoiceThread in Online Intermediate Spanish Language Classes*, The Coastal Review: An Online Peer-reviewed Journal: Vol. 13: Issue 1, Article 3, 1-16, DOI: 10.20429/cr.2023.130103
- Bezemer, Kress 2015: J. Bezemer, & G. Kress, *Multimodality, learning and communication: A social semiotic frame*, London, Routledge, <https://doi.org/10.4324/9781315687537>
- Bodis et al. 2020: A. Bodis, M. Reed, Y. Kharchenko, *Microteaching in Isolation: Fostering Autonomy and Learner Engagement through VoiceThread*, International Journal of TESOL Studies, Vol. 2 (3), 1-12, <https://doi.org/10.46451/ijts.2020.09.14>
- Chen L. 2025: L. Chen, *Exploring the Use of VoiceThread in a Peer-Mentored EMI Classroom: A Q-Methodology Protocol*, International Journal of Qualitative Methods, Volume 24, 1-10, DOI: 10.1177/16094069251326413
- Chen M. 2025: M. Chen, *Out-of-class support for the flipped language classroom: using VoiceThread microlectures to boost active teaching and learning*, System 132 103723, 1-20, <https://doi.org/10.1016/j.system.2025.103723>
- deNoyelles et al. 2014: A. deNoyelles, J. Zydney, B. Chen, *Strategies for creating a Community of Inquiry through online asynchronous discussions*, MERLOT Journal of Online Learning and Teaching, Vol. 10, No. 1, 153–165.
- Diamantopoulou, Ørevik 2025: S. Diamantopoulou, & S. Ørevik, *Multimodality in English Language Learning*, In: Chapelle, C, (ed.) Wiley Encyclopedia of Applied Linguistics, Wiley: Hoboken, NJ, USA, <https://doi.org/10.1002/9781405198431.wbeal20822>
- Donaghy, K. (2024). *The Multimodal Approach in Language Teaching*, Proceedings, 37th Annual APPI Conference, retrieved from: https://www.appi.pt/storage/app/media/Mindful%20EFL%20SIG/Kieran%20Donaghy_the-multimodal-approach-in-language-teaching.pdf, 11.02.2026.
- Garrison et al. 1999, D. R. Garrison, T. Anderson, W. Archer, *Critical inquiry in a text-based environment: Computer conferencing in higher education*, The Internet and Higher Education, Volume 2, Issues 2–3, 87-105, [https://doi.org/10.1016/S1096-7516\(00\)00016-6](https://doi.org/10.1016/S1096-7516(00)00016-6)
- Hendrawaty et al. 2024: N. Hendrawaty, Z. Sakhiyya, S. Wahyuni, Yuliati, *The Multimodal Approach in English Language Teaching: A Systematic Review*, Proceedings of Fine Arts, Literature, Language, and Education, 477-492
- Krsmanović et al. 2025: I. Krsmanović, D. Archan, L. Tica, *Teaching Presence for Inclusiveness in Blended Learning Classrooms: Perspectives from English Language Educators*, Computer Assisted Language Learning Electronic Journal (CALL-EJ), 26(2), 83-117
- Krsmanović 2021: Engineering students' attitudes towards Computer-assisted language learning in EFL, *Belgrade English Language and Literature Studies*, 2021, 13 (1), 93-123. <https://belgrade-bells.fil.bg.ac.rs/index.php/belgrade-bells/article/view/bells-2021-13-4>
- Lim et al. 2022: Lim, F. V., Toh, W., & Nguyen, T. T. H. (2022). Multimodality in the English language classroom: A systematic review of literature. *Linguistics and Education*, 69, 101048. <https://doi.org/10.1016/j.linged.2022.101048>
- Utami, Santosa 2023: N. L. G. F. P. Utami, & M. H. Santosa, *Using VoiceThread to develop the speaking skills of EFL learners: A systematic literature review*, Jurnal Inovasi Teknologi

Pembelajaran: Kajian dan Riset Dalam Teknologi Pembelajaran, Vol 10, No 1, 83-91, <https://doi.org/10.17977/um031v10i12023p083>

Ward et al. 2017: Y. D. Ward, J. G. Ward, M. (E.) Tao, *Using VoiceThread to Promote Active Learning and Learner Engagement in an Online Business Course*, Fort Hays State University, <https://scholars.fhsu.edu/cgi/viewcontent.cgi?article=1024&context=sacad>, 10.02.2026.

Zemlyanova et al. 2021: Zemlyanova, M., Muravyeva, N., Masterskikh, S., Shilova, L., & Shevtsova, A. (2021). Advancing English Language Learners' Speaking Skills Using VoiceThread in Mobile Learning for Russian Tertiary Context. *International Journal of Web-Based Learning and Teaching Technologies (IJWLTT)*, 16(6), 1-11.

Ivana M. Krsmanović, Marija D. Nikolić / ПРИМЕНА МУЛТИМОДАЛНИХ ДИГИТАЛНИХ АЛАТА У ESP УЧИОНИЦИ: ПЕДАГОШКИ ПОТЕНЦИЈАЛ ПЛАТФОРМЕ VOICETHREAD

Резиме / Рад истражује педагошки потенцијал платформе *VoiceThread*, мултимодалног дигиталног алата, у оквиру рачунарски подржаног учења језика (CALL) у настави енглеског језика струке (ESP). Полазећи од све израженије потребе за развојем професионално усмерене комуникативне компетенције студената, у раду је приказан праксом утемељен, дескриптивно-аналитички приступ, који има за циљ да испита могућности интеграције мултимодалних дигиталних алата у наставу енглеског језика струке, као и њихов допринос развоју говорних вештина и професионалног дискурса у ESP контексту. Рад илуструје како се *VoiceThread* може систематично интегрисати у ESP учионицу ради подршке говорним активностима специфичним за одређене дисциплине, кроз асинхрону, мултимодалну интеракцију. Ослањајући се на посматрање активности кроз наставну праксу, рад илуструје принципе дизајнирања задатака и улогу наставничке медијације у контекстима ESP наставе подржане CALL-ом. Опсервациони увиди указују на више педагошких предности, укључујући повећану ангажованост студената, већу спремност за учешће у говорним активностима, унапређено познавање конвенција професионалног дискурса, као и могућности за рефлексiju и вршњачку интеракцију изван ограничења учионице. Рад закључује да мултимодални CALL алати попут *VoiceThread*-а, могу функционисати као ефикасно педагошка подршка у настави енглеског језика струке.

Кључне речи: Енглески језик струке (ESP), *VoiceThread*, мултимодалност, рачунарски подржано учење језика (CALL), дигитална педагогија.

Примљен: 12. 2. 2026.

Прихваћен марта 2026.