

Beyond the Textbook: A Conceptual Framework for Integrating Multimedia Pedagogy in Technical English Classrooms

¹Vidya Sagar Rao Podicheti , ²Dr. Maheepathi Srinivasa Rao

¹Lecturer in English Government Polytechnic, Nalgonda

²Associate Professor of English Anurag University, Hyderabad

Abstract: English Language Teaching in polytechnic institutes across India has long been confined to textbook-driven instruction that develops grammatical knowledge but fails to build the oral communication skills that industry demands. This paper proposes the Triple-I Framework, a three-stage conceptual model comprising Immersion, Interaction, and Integration, designed to shift polytechnic ELT from passive text-based learning to active, multimedia-supported oral production. Grounded in Mayer's Cognitive Theory of Multimedia Learning, Krashen's Affective Filter Hypothesis, Vygotsky's Zone of Proximal Development, and Swain's Output Hypothesis, the framework uses authentic video input, digitally scaffolded practice through shadowing and self-recording, and real-time peer interaction to develop industry-ready spoken English. A job interview preparation module illustrates the framework's classroom application across three structured sessions. The paper also situates the framework within the NEP 2020 mandate for skill-embedded, vocationally relevant education. As a conceptual model, the Triple-I Framework offers polytechnic English teachers a theoretically grounded, practically viable, and infrastructure-light pathway for developing the oral proficiency that diploma students need to succeed in professional recruitment contexts.

Keywords: Multimedia Pedagogy, Triple-I Framework, Oral Communication, Polytechnic Education, Technical English, Digital Scaffolding, Task-Based Learning, NEP 2020, Employability, ELT.

Introduction

English communication skills have become as critical as technical expertise in today's workplace. Yet in most polytechnic institutes across India, English instruction remains confined to textbooks that prioritize grammar and reading over speaking and listening. The result is a widening gap between what the classroom produces and what the industry demands.

The scale of this gap is well documented. Research confirms that communication skills among polytechnic students are considerably low, directly affecting their employability and confidence (Arulselvi, 2018). Many students enter the job market with strong technical knowledge but are unable to introduce themselves fluently to an interviewer or participate effectively in a workplace discussion. They can complete a grammar exercise but cannot reproduce that knowledge under the pressure of real communication. This gap between passive textbook knowledge and active oral performance is the central problem this paper addresses.

Solving it requires more than minor curriculum adjustments. It requires rethinking how English is taught, in a way that aligns with the hands-on learning styles of vocational students, the communicative demands of industry, and the vision of the National Education Policy (NEP) 2020, which calls for English to function as a practical vocational skill rather than an isolated academic subject.

This paper proposes that multimedia pedagogy offers the most viable pathway for making that shift. Combining audio, video, and task-based digital activities can transform the English classroom from a site of passive reception



into a space for authentic oral production, while also lowering the anxiety that prevents many polytechnic students from speaking confidently in English.

The paper pursues three objectives: to examine the theoretical foundations supporting multimedia in oral ELT; to identify the limitations of textbook-driven pedagogy in the polytechnic context; and to propose the Triple-I Framework, a three-stage instructional model of Immersion, Interaction, and Integration, illustrated through a job interview preparation module. The Literature Review, Theoretical Foundations, Framework, Case Illustration, and Challenges sections that follow develop each of these in turn.

Literature Review

The disconnect between classroom English and workplace communication is well documented in Indian technical education. Research on polytechnic students confirms that communication skills in this population are considerably low, directly impacting employability and professional confidence (Arulselvi, 2018). Clement (2015) found that even final-year engineering students lack the confidence to face campus placements and expressed a clear preference for interactive language practice over conventional instruction. Dutte (2024) further confirms that verbal communication and presentation ability significantly influence hiring decisions, while technical curricula continue to neglect soft skills development. The problem, as these studies collectively indicate, is not student aptitude but pedagogical method: text-heavy, grammar-focused instruction does not develop the spoken competence that industry demands.

The inadequacy of monomodal instruction for oral skill development is equally well established. Richards (2006) argues that communicative language teaching requires genuine interaction, not scripted textbook exercises. Mayer (2001) demonstrates that learners process and retain language more effectively when auditory and visual channels are engaged simultaneously, a condition that printed text cannot meet. Building on this, Fahri et al. found significant improvement in fluency, accuracy, and speaking confidence when multimedia tools including video modeling and AI-based applications were integrated into a structured instructional sequence. In vocational college contexts specifically, multimedia teaching models have been shown to promote learner autonomy and improve overall English performance (Zhang, 2023).

Task-based approaches add the dimension of output. Lume and Hisbullah (2022) show that learners in task-based instruction consistently outperform those in conventional grammar-driven classes on oral performance measures. When digital scaffolding is integrated into task-based learning, gains are stronger: Fang et al. (2021) found that mobile-supported task-based instruction with built-in linguistic scaffolds produced greater vocabulary and oral production outcomes than paper-based task-based learning alone. Swain's (1985) Output Hypothesis provides the theoretical explanation: learners need pushed output, not just comprehensible input, to notice gaps in their language and develop the automaticity that fluent speech requires. Vygotsky's (1978) Zone of Proximal Development completes the picture by explaining how graduated scaffolding, progressively withdrawn as competence grows, moves learners from dependence to independent oral performance.

Taken together, the literature establishes that multimedia input, task-based output, and scaffolded progression are the three conditions most consistently associated with oral proficiency gains in vocational learners. What remains underdeveloped is a structured pedagogical model that integrates all three conditions specifically for the Indian polytechnic classroom and aligns with the NEP 2020 mandate for skill-embedded, industry-relevant English instruction. The Triple-I Framework proposed in this paper addresses that gap.



Theoretical Foundations

Four established frameworks from cognitive psychology and second language acquisition collectively justify the multimedia approach this paper proposes.

Mayer's (2001) Cognitive Theory of Multimedia Learning establishes that the human mind processes information through two parallel channels, auditory and visual, and that learning is significantly more effective when both are engaged simultaneously. For oral communication teaching, this means that video-based input, which carries voice, gesture, facial expression, and situational context together, is cognitively superior to printed text, which carries only one of these at a time.

Krashen's (1982) Affective Filter Hypothesis explains why environment matters as much as content. When anxiety is high, language acquisition is blocked regardless of instructional quality. Polytechnic students preparing for job interviews face precisely this kind of anxiety around English. Digital tools such as recording apps and AI simulators create a private, low-stakes practice space that lowers this filter and makes acquisition possible before students face real human interaction.

Vygotsky's (1978) Zone of Proximal Development provides the scaffolding rationale for the framework's three-phase design. Learners advance most efficiently when support is calibrated to their current level and gradually withdrawn as competence grows. Each phase of the Triple-I Framework represents a different level of scaffolding, maximal in Phase I, mediated in Phase II, and absent in Phase III, mirroring this principle directly.

Swain's (1985) Output Hypothesis adds the final and most practically significant argument: comprehensible input alone is not sufficient for oral fluency. Learners must be pushed to produce output in order to notice gaps in their language, test new forms, and develop the automaticity that spontaneous speech requires. This is why the framework moves beyond observation and input toward recorded self-practice and live unscripted performance.

Together, these four theories describe a complete instructional logic. Mayer justifies multi-modal input. Krashen justifies the low-stakes digital environment. Vygotsky justifies the phased, scaffolded progression. Swain justifies the non-negotiable demand for output. The Triple-I Framework is built at the intersection of all four.

The Triple-I Conceptual Framework

The Triple-I Framework proposes a three-stage instructional sequence that moves the polytechnic learner progressively from supported observation to controlled practice to independent oral production. The stages are Immersion, Interaction, and Integration, and they must be followed in order: each phase builds on the one before it.

Phase I: Immersion (Sensory Input)

The textbook is replaced by authentic or near-authentic multimedia input. Students watch curated videos of professional interactions, job interviews, workplace briefings, and technical presentations, using a guided observation checklist to analyse both verbal features such as vocabulary, register, and sentence structure, and non-verbal features such as eye contact, posture, and pacing. Students do not speak in this phase. The goal is perceptual readiness: a clear, multi-modal understanding of what competent professional communication looks and sounds like, grounded in Mayer's (2001) Contiguity and Modality Principles.

Phase II: Interaction (Controlled Practice)

Students move from observation to production in a low-stakes digital environment. Using the shadowing technique, they listen to short professional clips and attempt to reproduce the speaker's intonation, stress, rhythm, and register. Research across 44 studies confirms that shadowing improves comprehensibility, fluency, and prosodic control in second language learners (Foote & McDonough,). Students then record their own responses to structured prompts, play back the recordings, identify fillers and pacing problems using a self-evaluation rubric, and re-record. This cycle of produce, listen, reflect, and revise activates Swain's (1985) noticing and hypothesis-testing functions and lowers the affective filter through the absence of a social audience (Krashen, 1982). Tools for this phase include mobile recording apps, AI interview simulators, and language lab software.

Phase III: Integration (Real-World Output)

The digital scaffolding is withdrawn. Students conduct peer-to-peer mock interviews via video conferencing, simulating the format of a remote recruitment interview. One student plays the HR Manager, the other the Diploma Candidate, and roles are reversed after each round. Questions are partly familiar from Phase II and partly new, requiring spontaneous, unscripted response. A peer assessment rubric covering clarity, confidence, vocabulary, and non-verbal communication provides structured feedback. This phase operationalizes Vygotsky's (1978) principle that scaffolding must eventually be faded to produce genuine independence.

The table below summarizes the framework:

Phase	Theoretical Basis	Core Activity	Key Tools	Outcome
I: Immersion	Mayer (2001)	Guided video observation	Authentic videos, observation checklist	Perceptual readiness
II: Interaction	Krashen (1982), Vygotsky (1978), Swain (1985)	Shadowing, recording, self-review	Recording apps, AI simulators	Controlled oral competence
III: Integration	Vygotsky (1978)	Peer mock interviews	Video conferencing, peer rubrics	Independent oral performance

The framework is a conceptual model, not a tested intervention. Its sequence is non-negotiable but its duration per phase is flexible, depending on cohort proficiency, available infrastructure, and semester schedule. Phase II activities can be completed on basic smartphones without a data connection, making the framework viable even in resource-constrained institutions.

Pedagogical Case Illustration: The Job Interview Module

To demonstrate the Triple-I Framework in practice, this section outlines a three-session module for first-year polytechnic students targeting a single high-stakes oral task: the job interview.

Objective: Students should be able to introduce themselves professionally, describe their technical skills and diploma project, and respond to common HR questions with appropriate register and confidence.

Session 1: Immersion. The instructor plays two contrasting interview videos. One candidate has strong technical knowledge but poor communicative competence: mumbling, no eye contact, and an informal register. The other



communicates with confidence and structure. Students use a guided observation checklist to identify specific verbal and non-verbal markers in each clip, then discuss what distinguished the second candidate.

Session 2: Interaction. Students shadow a model self-introduction clip, matching the speaker's intonation, rhythm, and register. They then record their own response to the prompt "Introduce yourself and tell me about your diploma project," listen to the playback, complete a self-evaluation rubric identifying fillers and pacing issues, and record a revised version.

Session 3: Integration. Students conduct peer mock interviews in pairs, one playing the HR Manager using a structured question bank, the other the Diploma Candidate. Roles are reversed after ten minutes. A peer assessment rubric covering confidence, clarity, vocabulary, and response structure guides the feedback.

Assessment: An Audio-Visual Portfolio comprising the student's best recording from Session 2 and the completed peer rubric from Session 3 replaces the conventional written test, directly measuring the oral competency the module was designed to build.

Policy Alignment: NEP 2020 and Vocational Excellence

The National Education Policy 2020 calls for the seamless integration of vocational education into mainstream curricula, with a clear emphasis on practical skills, digital tools, and industry-relevant learning (Ministry of Education, 2020). Within this policy frame, English in a polytechnic cannot remain an isolated grammar subject. It must function as a vocational communication skill.

The Triple-I Framework aligns directly with this vision. By grounding English instruction in authentic professional contexts, using digital tools as pedagogical infrastructure, and assessing students through oral performance rather than written examinations, the framework treats spoken English as what NEP 2020 implicitly requires it to be: a workplace-ready skill, not an academic exercise. The Audio-Visual Portfolio proposed in the case illustration is one concrete step toward the kind of skill-based, continuous internal assessment that NEP 2020 recommends as an alternative to high-stakes written testing.

Challenges and Strategic Solutions

The Digital Divide. Students from rural backgrounds may have limited access to smartphones or reliable internet. Most Phase II activities, however, require only a basic recording app that functions offline. Institutions should prioritize at least one well-equipped communication lab to ensure equitable access for Phase III activities.

Curriculum Rigidity. Boards of Technical Education typically mandate written examinations as the primary assessment instrument. Teachers and institutional heads should advocate for Audio-Visual Portfolios to be recognized as valid internal assessment components, a step that is consistent with NEP 2020's push toward continuous and skill-based evaluation.

Teacher Readiness. Many polytechnic English teachers are trained as grammarians rather than communication facilitators. Transitioning to the Triple-I Framework requires teachers to curate digital content, operate basic recording and video conferencing tools, and design rubrics for oral assessment. Short, focused professional development workshops, ideally organized at the State Board level, are essential to make this transition sustainable.

Student Resistance. Students accustomed to passive, textbook-driven learning may initially resist being asked to record themselves or speak in front of peers. Beginning Phase II with fully private, unshared recordings and introducing peer interaction gradually can ease this resistance without compromising the framework's integrity.



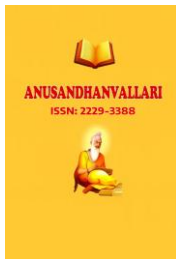
Conclusion

Traditional ELT in polytechnic institutes has long prioritized grammatical accuracy over communicative competence, producing graduates who are technically qualified but orally underprepared for the workplace. This paper has argued that multimedia pedagogy, structured through the Triple-I Framework of Immersion, Interaction, and Integration, offers a theoretically grounded and practically viable alternative. By combining multi-modal input, digitally scaffolded practice, and real-time unscripted output, the framework addresses the specific oral communication needs of vocational learners in a way that traditional textbooks cannot.

The Triple-I Framework is expected to produce three outcomes. First, it lowers the affective barrier to spoken English by giving students private, low-stakes digital practice before demanding public performance. Second, it develops perceptual awareness of professional communication norms through authentic video input, equipping students with a concrete model to work toward. Third, it builds transferable oral proficiency through progressive scaffolding that ends in genuine, unscripted human interaction. Taken together, these outcomes directly address the employability gap documented in the literature. For institutions, the framework's most significant implication is that meaningful oral communication training does not require expensive infrastructure. It requires a reorientation of purpose: from teaching about English to teaching through English, in contexts that mirror the professional world students are about to enter. As a conceptual model, the framework now awaits empirical validation through classroom-based comparative studies and longitudinal tracking of student performance in actual placement drives.

Works Cited

- [1] Arulselvi, E. (2018). Communication skills among polytechnic students. *Language in India*, 18(1), 45-52.
- [2] Chapelle, C. A. (2003). *English language learning and technology: Lectures on applied linguistics in the age of information and communication technology*. John Benjamins Publishing Company.
- [3] Clement, A. (2015). English for employability: A case study of the English language training need analysis for engineering students in India. *English Language Teaching*, 8(2), 116-125.
- [4] Dutte, S. S. (2024). Bridging the employability gap: The importance of English communication skills for engineering graduates. *International Journal of Advanced Research and Multidisciplinary Trends*, 1(3).
- [5] Kavuri, S. (2024). Shift-Left and Shift-Right Testing Approaches: A Practical Roadmap for Continuous Quality in Agile and DevOps. *Journal of Information Systems Engineering and Management*, . <https://doi.org/10.52783/jisem.v9i4.127>
- [6] Fang, X., Xu, J., & Mehta, P. (2021). Effects of mobile-supported task-based language teaching on EFL students' linguistic achievement and conversational interaction. *ReCALL*, 33(1), 1-17.
- [7] Gummadi, V. P. K. (2023). MuleSoft batch processing: High-volume streaming architecture. *Computer Fraud & Security*, 2023(12), 50–57. <https://doi.org/10.52710/cfs.886>.
- [8] Hamada, Y. (2018). Shadowing: Who does it work for? *Language Teaching Research*, 23(4), 386-406.
- [9] Hockly, N. (2013). *Mobile learning*. Oxford University Press.
- [10] Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.



-
- [11] Lume, L. L., & Hisbullah, M. (2022). The effectiveness of task-based language teaching to teach speaking skills. *Journal of Languages and Language Teaching*, 10(1), 85-93.
 - [12] Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.
 - [13] Ministry of Education. (2020). *National Education Policy 2020*. Government of India.
 - [14] Richards, J. C. (2006). *Communicative language teaching today*. Cambridge University Press.
 - [15] Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Newbury House.
 - [16] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
 - [17] Zhang, W. (2023). Using multimedia information retrieval technology to optimize the strategies of college English teaching. *Journal of Educational Technology*, 12(2), 45-58.