



Improving English Listening Skills for Vocational Students Through Interactive ICT Tools

Muhammad Husaein Jaya Negara¹, Jamilah²

^{1,2} Pendidikan Bahasa Inggris, Universitas Negeri Yogyakarta, DI Yogyakarta

Corresponding E-Mail: muhammadhusaein.2023@student.uny.ac.id

Received: 2025-10-06 Accepted: 2025-10-17

DOI: 10.24256/ideas.v13i2.8133

Abstract

Vocational students often face difficulties in developing English listening skills, particularly with authentic speech that includes varied accents, fast delivery, and technical terms. This Classroom Action Research conducted at SMK N 2 Depok explored the use of interactive ICT tools—YouTube and ESLVideo.com—to enhance listening comprehension among 36 eleventh-grade Computer and Network Engineering students. The intervention applied a Genre-Based Approach in two cycles, using authentic YouTube videos for exposure to real-world discourse and ESLVideo.com quizzes for interactive practice with immediate feedback. Data from tests, observations, and interviews showed substantial progress: students' mean listening scores increased from 71.25 to 90.83, accompanied by higher engagement and confidence. Qualitative findings revealed that visual and verbal integration reduced comprehension difficulty and motivated students to participate more actively. Despite challenges such as technical issues and classroom management, reflective adjustments helped maintain consistent improvement. The study concludes that integrating interactive ICT tools can effectively enhance vocational students' listening competence and foster independent learning aligned with professional communication needs.

Keywords: Listening skills; Vocational education; Interactive ICT tools.

Introduction

Listening comprehension is a fundamental component of second language acquisition, forming the foundation upon which speaking, reading, and writing skills develop (Goh & Vandergrift, 2021). As the primary channel for linguistic input, listening enables learners to engage in meaningful communication and internalize linguistic forms across academic and professional contexts (Field, 2008; Brown & Lee, 2015).

Despite its importance, listening remains one of the most demanding skills for

English as a Foreign Language (EFL) learners. Authentic spoken discourse often involves rapid delivery, varied accents, and idiomatic expressions that challenge learners unfamiliar with natural speech (Adi et al., 2022). Limited exposure to authentic materials and the absence of visual cues further increases cognitive demands and hinder comprehension efficiency (Goh & Vandergrift, 2021).

In Indonesia's vocational education system, the significance of listening is amplified by workplace communication demands. Vocational students must acquire communicative competence relevant to fields such as tourism, hospitality, and information technology, where English frequently serves as the medium of interaction (Dudley-Evans & St. John, 1998; Lertchalermtipakoon et al., 2021). Yet, classroom observations reveal persistent difficulties with authentic materials containing technical vocabulary and natural speech rates (Hadist et al., 2022).

Traditional instructional practices in vocational settings often rely on scripted textbook recordings and teacher-centered explanations. These methods limit exposure to spontaneous communication and reduce opportunities for active engagement, resulting in slow listening development and poor transfer to real-world contexts (Siregar, 2024).

From a cognitive standpoint, these difficulties can be explained through Cognitive Load Theory (Sweller, 1994), which posits that learners' working memory is easily overloaded when processing complex auditory input without adequate support. Dual Coding Theory (Paivio, 1986) and Mayer's Cognitive Theory of Multimedia Learning (2009) further suggest that comprehension improves when auditory and visual information are processed together. Integrating multimodal input—such as visuals, subtitles, and nonverbal cues—helps distribute cognitive effort and enhance retention.

From a motivational perspective, learners' engagement depends on the perceived relevance of tasks. When listening activities seem monotonous or disconnected from professional goals, intrinsic motivation declines. Conversely, authentic materials aligned with learners' vocational contexts can increase engagement and persistence (Gardner & Lambert, 1972; Husnawati et al., 2024). Effective pedagogy must therefore manage cognitive load and maintain motivational relevance.

Information and Communication Technology (ICT) provides promising tools to meet these dual challenges. Platforms such as YouTube and ESLVideo.com expose learners to authentic speech, diverse accents, and contextual cues that mirror real-world communication (Verbeke & Simon, 2023). These tools enable learners to control playback, repeat difficult sections, and receive immediate feedback, fostering autonomy and interactive learning (Pei et al., 2023).

For vocational learners, ICT-based materials offer distinct advantages: they introduce industry-specific discourse, reduce cognitive strain through synchronized visual-auditory input, and promote engagement through contextual relevance (Setiawan et al., 2023; Al-Jumaily & Alazzawi, 2025). Such integration

aligns with English for Specific Purposes (ESP) principles, emphasizing language learning tailored to occupational realities (Dudley-Evans & St. John, 1998; Aziz & Anjaniputra, 2025).

However, research on ICT-mediated listening instruction for vocational learners remains limited. Most studies focus on tertiary or general EFL contexts where learners have stronger linguistic foundations and broader communicative goals. Recent studies on podcast and video-based interventions (Valentin, 2024; Efendi & Astutik, 2024; Nofiatun & Purwanto, 2024) show positive outcomes, yet few have systematically examined the combined use of YouTube and ESLVideo.com within iterative, classroom-based frameworks.

Unlike previous studies, this research is the first to explore the integration of these two ICT tools within a Genre-Based Approach (GBA) implemented through Classroom Action Research (CAR) in Indonesian vocational education. This approach bridges the gap between cognitive-motivational theory and practical classroom application. This study seeks to answer two main questions:

- 1) How can interactive ICT tools improve the English listening skills of vocational students?
- 2) What challenges arise during their implementation, and how can these be addressed?

Theoretically, this study contributes by operationalizing cognitive and motivational frameworks within ICT-supported listening pedagogy. Practically, it proposes a replicable instructional model that demonstrates how systematic ICT integration can enhance comprehension, motivation, and learner autonomy in vocational English contexts.

Method

Research Design

This study employed a Classroom Action Research (CAR) design to address persistent listening comprehension difficulties among vocational learners. CAR was chosen because it allows systematic, iterative cycles of planning, action, observation, and reflection, enabling teachers and researchers to refine teaching practices based on classroom evidence (Kemmis & McTaggart, 1988; Burns, 2010).

The research consisted of two complete cycles conducted over six meetings within a four-week period during the 2024/2025 academic year. Each cycle comprised three classroom meetings integrated into the school's regular English schedule. This cyclical process enabled continuous evaluation and adaptation of instructional strategies in response to classroom needs.

Participants and Setting

The study was conducted at SMK N 2 Depok, Yogyakarta, with 36 Grade XI students from the Computer and Network Engineering (*SIJA 1*) program. The participants represented a typical vocational EFL context characterized by intermediate proficiency and limited exposure to authentic spoken English.

Preliminary classroom observations revealed that the students struggled to process rapid speech, varied accents, and technical vocabulary. The collaborating English teacher participated in all stages of the CAR process to ensure contextual relevance and curricular alignment. The school's ICT facilities—including projectors, Wi-Fi access, and a computer laboratory—supported the integration of YouTube and ESLVideo.com activities throughout the intervention.

Research Procedures

Each cycle of the CAR followed the four stages proposed by Kemmis and McTaggart (1988): planning, action, observation, and reflection.

In Cycle I, instruction focused on familiarizing students with authentic English input related to computer hardware and setup. Students watched short YouTube tutorials such as *Setting Up a Desktop Computer* and *Installing a Wireless Router*. Each video was followed by corresponding ESLVideo.com comprehension quizzes, which required students to identify main ideas, technical terms, and procedural steps.

During the pre-listening stage, students predicted content based on screenshots and discussed vocabulary lists prepared in Canva presentations. The while-listening stage involved viewing the videos with subtitles and answering ESLVideo.com quizzes that provided instant feedback. The post-listening stage consisted of small-group discussions and collaborative summaries written on Google Docs.

Cycle II emphasized comprehension of longer and more complex authentic materials. Students listened to videos such as *Installing Windows 10 from a USB Drive* and *Troubleshooting Network Connections*. The focus shifted from understanding explicit information to interpreting sequencing, cause-effect relations, and conditional instructions.

Activities included identifying transition markers (e.g., *first, next, after that*), recognizing modal verbs used for advice (*should, must*), and analyzing intonation patterns in procedural discourse. Vocabulary and grammar exercises were embedded within listening tasks to promote integrated skill development.

Reflection meetings after each cycle identified teaching adjustments. Based on feedback, teachers modified scaffolding strategies—for instance, by increasing preview time and reducing video speed for lower-performing students.

Data Collection

Multiple instruments were employed to gather both quantitative and qualitative data, ensuring triangulation and comprehensive analysis.

1) Listening Comprehension Tests

Pre-tests and post-tests were administered before and after each cycle to measure changes in listening performance. The tests, adapted from the TOEFL ITP® framework, contained 20 multiple-choice items divided into two

sections, each based on a short YouTube video relevant to vocational contexts such as *Setting Up a Desktop Computer* and *Setting Up a Wireless Router*. Each section consisted of 10 questions assessing both micro-skills (e.g., detail identification, vocabulary recognition) and macro-skills (e.g., main ideas, inference, sequencing). For example:

"According to the speaker, what is the first action that should be taken when beginning the setup?"

- a) Connecting the monitor cable*
- b) Unpacking all components from the box*
- c) Attaching the keyboard and mouse*

Test materials were contextualized to vocational topics such as computer setup and troubleshooting. Validation procedures included peer review, contextual adaptation, and expert evaluation by two applied linguists who verified the content's relevance and communicative authenticity (Bachman, 1990).

Internal consistency was established through a pilot test with a similar class (N = 30), yielding a Cronbach's $\alpha = .86$, indicating high reliability.

2) Observation Checklists and Field Notes

Classroom observations were conducted during each meeting by the researcher and the collaborating teacher using structured checklists. Field notes documented engagement, interaction patterns, and implementation challenges. These records also guided reflection and revision at the end of each cycle.

3) Interviews

Semi-structured interviews were conducted with the English teacher and a subset of students before, during, and after each CAR cycle. The pre-cycle interviews explored initial listening difficulties; the mid-cycle interviews captured experiences and motivation during ICT-based learning; and the post-cycle interviews identified perceived improvements and pedagogical implications.

Data Collection Procedures

Quantitative and qualitative data were collected continuously across both CAR cycles. Listening test results were gathered before and after each cycle, while observations and interviews occurred throughout instruction. Reflection meetings between the teacher and researcher followed each session to assess progress and determine necessary revisions for subsequent lessons. To ensure clarity in qualitative reporting, data excerpts presented in the Results section are coded as follows: "*ET*" refers to the English teacher, "*S*" to student participants, and "*R*" to the researcher. These codes are used to attribute quotations and observation notes while maintaining participant confidentiality.

Data Analysis

Both quantitative and qualitative analyses were employed in a mixed-methods triangulation framework.

1) Quantitative Analysis

Quantitative data from pre- and post-tests were analyzed using descriptive statistics (mean scores, gain percentages) and paired-samples t-tests to determine significant differences between cycles. Improvement in mean scores across cycles was used to evaluate the effectiveness of the ICT-integrated intervention.

2) Qualitative Analysis

Qualitative data from observations, field notes, and interviews were analyzed through thematic coding. Codes were organized into three major categories: *comprehension improvement, motivation and engagement, and implementation challenges*. Coding reliability was verified through inter-rater agreement ($\kappa = 0.87$), indicating high consistency between coders.

3) Triangulation

Following Denzin's (2017) model, methodological triangulation was applied to cross-validate findings from multiple data sources. Quantitative results (test scores) were compared with qualitative evidence (observations, interviews, and field notes). For instance, gains in mean listening scores were matched with increases in observed participation and self-reported confidence.

Additionally, data triangulation was implemented by collecting perspectives from different participants (students and teacher) and at different time points (pre-, mid-, and post-cycle). The convergence of results across methods and sources enhanced the study's credibility and internal validity.

Ethical Considerations

Ethical approval was granted by the Research Ethics Committee of the English Education Department, Universitas Negeri Yogyakarta. Participation was voluntary, and informed consent was obtained from all students and school administrators. To maintain confidentiality, participants' names were replaced with coded identifiers. The intervention was designed to enhance, not disrupt, the existing curriculum, ensuring that all research activities contributed positively to the learning process.

Results***Improvement in Students' Listening Skills***

Before the intervention, students' listening proficiency was below the expected standard. The mean pre-test score was 71.25, notably lower than the school's mastery criterion of 80. Classroom observations during regular lessons showed that students often hesitated to answer comprehension questions, waited

for teacher translation, and rarely initiated discussion. The English teacher explained:

"Students understand grammar well but become confused when listening to natural English, especially when the speaker talks fast." —ET

During pre-cycle discussions, several learners shared similar difficulties. They described authentic audio as *"too fast and continuous"*, which made it difficult to distinguish words or grasp the overall meaning.

"Sometimes I only catch one or two words." —S

"If the speaker talks too fast, I just stop listening because I can't follow." —S

These insights revealed a gap between the scripted textbook recordings commonly used in class and the spontaneous, unsimplified speech in authentic contexts. The diagnostic data thus justified the need for an intervention that combined authentic multimedia input, visual support, and interactive engagement, which became the foundation of the YouTube and ESLVideo.com integration used in this study.

Cycle I Implementation

Cycle I introduced YouTube tutorial videos and ESLVideo.com quizzes within a Genre-Based Approach (GBA). Lessons followed three structured stages—pre-listening, while-listening, and post-listening—to scaffold comprehension systematically.

In the pre-listening stage, the teacher presented key vocabulary and technical terms relevant to the learners' field (e.g., *motherboard, router, network cable*). During while-listening, students viewed authentic tutorials such as *"Setting Up a Desktop Computer"* and *"Installing a Wireless Router"*. They predicted procedural steps and answered comprehension questions via ESLVideo.com quizzes, receiving instant feedback.

The first meetings revealed several challenges. Many students struggled to keep up with rapid speech and unfamiliar accents, frequently replaying video segments. Technical interruptions like buffering sometimes broke concentration.

"When the video buffers, I forget what I just heard." —S

Despite these initial setbacks, improvement began to emerge by the third meeting. Students gradually recognized organizational patterns in procedural texts and began connecting auditory input with visual demonstrations.

"Oh, so the goal comes first, then the materials and steps." —S

This remark illustrated growing awareness of discourse structure and procedural sequencing. The combination of visual and verbal input helped reduce

cognitive load, allowing students to process meaning through multiple channels.

The end-of-cycle test results confirmed the observed progress. The mean score increased from 71.25 (Pre-Test) to 83.19 (Post-Test 1). This substantial gain showed that authentic, multimodal materials effectively enhanced comprehension.

This improvement directly addresses the first research question by demonstrating that interactive ICT tools can significantly enhance students' listening comprehension when integrated through structured, genre-based instruction. The combination of visual and auditory input from YouTube and the immediate feedback from ESLVideo.com supported active meaning-making, reduced cognitive load, and promoted sustained learner engagement. These outcomes suggest that authentic, multimodal exposure helps vocational learners bridge the gap between textbook-based practice and real-world English communication.

During reflection, the teacher noted positive behavioral changes:

"Students who were usually quiet began to answer voluntarily, especially when the video matched something they had done in computer practice." —ET

These observations suggested that connecting listening content to vocational experience not only facilitated comprehension but also boosted learner motivation.

Cycle II Implementation

Cycle II refined the teaching process by addressing technical and linguistic challenges from Cycle I. All videos were pre-downloaded, devices were checked before class, and vocabulary lists were expanded to include recurring technical expressions. Brief pronunciation and prediction activities were added to prepare students for authentic listening.

Students engaged with more complex videos such as *"Installing Windows 10 from a USB Drive"* and *"Troubleshooting Network Errors."* By this point, they demonstrated stronger control of sequence markers and imperative verbs and were more capable of following procedural discourse.

"It says 'click here' and then 'open settings'—that shows the order." —S

"We must use 'then' or 'after that' so the steps are clear." —S

These utterances illustrated deeper understanding of discourse organization and the language of instruction. In a follow-up transcript-analysis activity, one student critically commented on video clarity:

"This part is unclear—it just says 'do the next step,' but doesn't tell which one." —S

Such remarks indicated that learners were not only understanding but also analyzing content quality—reflecting a transition from comprehension to critical

awareness, characteristic of higher listening proficiency.

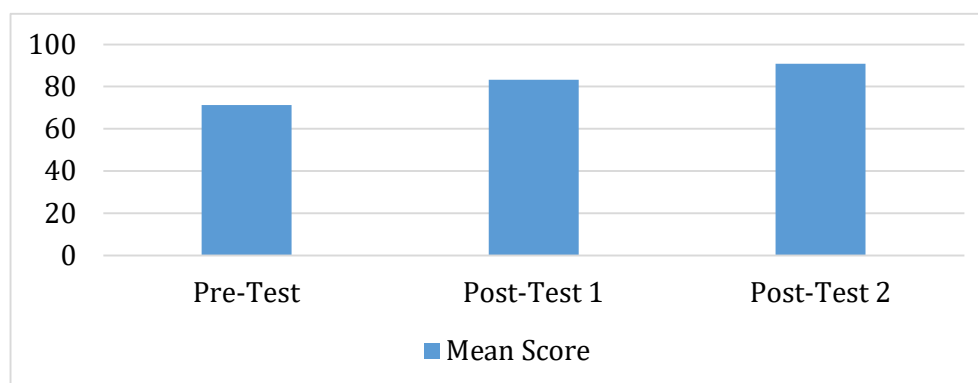
By the end of Cycle II, the mean score reached 90.83 (Post-Test 2), an additional 7.64-point increase from Cycle I. The cumulative gain of 19.58 points from pre-test to Post-Test 2 indicated steady and meaningful progress across the two CAR cycles (see Table 1 and Figure 1).

The results from Cycle II provide further evidence of the intervention's effectiveness and connect to the second research question concerning implementation challenges. The observed progress indicates that reflective adjustment—such as pre-downloading materials and expanding vocabulary scaffolding—played a crucial role in sustaining improvement. This cycle confirmed that systematic reflection and adaptation are essential components of ICT-based listening instruction, allowing teachers to fine-tune support mechanisms while maintaining learner autonomy.

Table 1. Summary of Listening Test Results

Test	N	Min	Max	Mean	SD	Mean Difference	t(35)	p	Cohen's d
Pre-Test	36	60	80	71.25	5.78	–	–	–	–
Post-Test 1	36	75	90	83.19	3.99	+11.94	-10.10	<.001	1.68
Post-Test 2	36	85	95	90.83	3.49	+7.64	-9.74	<.001	1.62

Figure 1. Mean Score Progression Across Tests



As illustrated in Figure 1, students' mean listening scores improved steadily across the three assessment points, reflecting both cognitive and motivational growth. The large effect sizes (Cohen's $d > 1.6$) highlight the practical significance of these gains, confirming that the integration of ICT tools led not only to statistical improvement but also to meaningful enhancement in learners' comprehension abilities and confidence.

The upward trajectory in mean scores across both cycles mirrors the

behavioral and attitudinal improvements observed in class. Students became more confident and proactive, often assisting peers and seeking clarification independently.

"They became more confident and tried to infer meaning before asking for translation." —ET

"Watching and listening together makes it easier to understand and more enjoyable than just hearing audio." —S

These experiences point to three interrelated processes underlying improvement:

- 1) Cognitive scaffolding through visual-verbal integration that reduced comprehension difficulty.
- 2) Contextual relevance that increased intrinsic motivation and engagement.
- 3) Interactive feedback that nurtured autonomy and reflective learning.

The results collectively affirm the iterative value of Classroom Action Research (CAR)—continuous reflection and adjustment leading to significant gains in both comprehension and learner engagement.

Challenges in Implementation and Iterative Responses

Although overall outcomes were positive, several key challenges emerged during implementation. Technical instability, such as buffering and unstable connections, initially disrupted lesson flow. Linguistic and cognitive overload also affected early sessions as students adjusted to authentic, fast-paced speech. Motivation and digital literacy presented additional barriers, with some learners finding online platforms difficult to navigate at first.

These challenges were gradually mitigated through reflective teaching actions, including pre-downloading videos, targeted vocabulary pre-teaching, and peer mentoring for ICT navigation. By Cycle II, most students demonstrated improved confidence and smoother participation. These findings emphasize that effective ICT integration requires not only appropriate tools but also careful pedagogical planning, technical preparedness, and continuous reflection to ensure equitable and engaging learning experiences.

Overall, these results confirm that structured ICT integration, supported by reflective teaching, can lead to substantial and sustainable improvement in vocational students' listening skills.

Discussion

This study investigated how interactive ICT tools—specifically YouTube and ESLVideo.com—can enhance vocational students' listening comprehension and what challenges arise during their implementation. The results demonstrate that authentic, multimodal, and interactive materials can significantly improve comprehension and motivation when systematically embedded within a Genre-Based Approach (GBA) and implemented through Classroom Action Research (CAR)

cycles.

The continuous improvement in listening scores across both cycles aligns with recent findings that ICT-supported listening fosters exposure to authentic input and contextual learning opportunities (Shamsi & Bozorgian, 2024; Khairat, 2024). However, the present study extends these findings by emphasizing that technology alone does not drive improvement—pedagogical design and structured scaffolding are crucial mediators. The use of pre-listening vocabulary activation, guided prediction, and reflective post-listening tasks enabled learners to internalize linguistic and discourse patterns rather than passively consuming multimedia content.

The visual dimension of YouTube videos effectively reduced cognitive strain by distributing information processing across auditory and visual channels, supporting the Dual Coding Theory (Paivio, 1986) and Cognitive Theory of Multimedia Learning (Mayer, 2009). Yet this study demonstrates that multimodal benefits materialize most effectively when combined with pedagogical scaffolding that guides attention, interpretation, and reflection. The increased comprehension observed in Cycle II aligns with recent findings that authentic multimedia input paired with structured instructional design leads to gradual and sustained understanding (Shamsi & Bozorgian, 2024).

Moreover, students' transformation from passive listeners to self-regulated learners reflects growth in metacognitive awareness—the ability to monitor, evaluate, and manage comprehension. This finding supports Goh and Vandergrift's (2021) model of metacognitive listening instruction, which views comprehension as an active, cyclical process of prediction, verification, and reflection. By the end of Cycle II, learners not only understood procedural content but also critiqued unclear instructional steps, indicating developing reflective and analytical listening behavior.

Finally, motivational changes emerged as a significant outcome. Initially, many students perceived listening tasks as tedious and overly challenging—especially when materials were static, speed was fast, or practice was limited—consistent with findings in recent Indonesian EFL studies (Febriani et al., 2022). As interactive ICT tools were introduced, however, students reported higher engagement, enjoyment, and confidence. This shift aligns with Self-Determination Theory (Deci & Ryan, 2000) and recent empirical findings indicating that learning environments promoting autonomy, competence, and relatedness foster sustained motivation. For example, Shao et al. (2025) found that digital technology use that allows learner control, feedback, and engaging interaction significantly boosts learning engagement, and Zhang & Crawford (2024) showed that gamified assessments provide feedback and choice—helping meet psychological needs and deepen engagement.

Theoretical Implications

The findings contribute to refining several theoretical models of multimedia

learning and language acquisition. They reinforce Cognitive Load Theory (Sweller, 1994) by illustrating how multimodal input—when pedagogically structured—reduces extraneous load and supports working memory processing. The study also adds nuance to Mayer’s (2009) multimedia learning framework, showing that pedagogical scaffolding and contextual relevance function as moderating variables that determine the success of multimedia integration.

Additionally, the results support Krashen’s (1982) Input Hypothesis by demonstrating that authentic multimedia exposure provides comprehensible input when appropriately scaffolded. The GBA stages operationalized this process, helping learners progress from understanding surface-level meaning to analyzing structure and function. This aligns with Field’s (2008) view that effective listening pedagogy must move beyond comprehension checking toward strategy development and reflection (Nasution & Prihatini, 2024).

The iterative CAR design further advances teacher reflection theory (Burns, 2010; Kemmis, McTaggart, & Nixon, 2014), illustrating how cyclical inquiry can serve as a mechanism for continuous improvement in ICT-based instruction. The teacher’s reflective adjustments—from addressing buffering issues to refining scaffolding—embody Schön’s (1983) concept of the “reflective practitioner.” This dynamic relationship between practice and theory supports the view that successful CALL outcomes depend more on pedagogical intentionality than on technology itself (Liu et al., 2023).

Although external factors such as novelty effects (Warschauer, 2011) or practice familiarity may have influenced initial enthusiasm, the sustained gains in learner autonomy, strategy use, and motivation observed (e.g. in studies with delayed post-testing) indicate genuine cognitive and affective development (Zhang & MacWhinney, 2023). Thus, this study contributes to a growing body of evidence suggesting that reflective, scaffolded ICT integration promotes not only linguistic competence but also self-regulated learning—an outcome supported by recent educational design studies and interventions explicitly linking technology use with SRL components (Zuo et al., 2024; Pansri et al., 2025).

Pedagogical Implications

Pedagogically, the findings underline that ICT tools achieve the greatest impact when integrated into structured, reflective instruction rather than employed as supplementary media. Teachers in vocational contexts should select authentic, field-specific multimedia materials that directly connect English input to professional tasks. This contextual alignment—demonstrated in recent ESP developments in Indonesia (Lave & Wenger, 1991; Santika et al., 2022)—ensures relevance and facilitates situated learning, allowing learners to process English as a tool for technical communication rather than as an abstract subject.

Equally crucial is the implementation of explicit scaffolding techniques. Teachers should pre-teach essential vocabulary, encourage guided prediction

before listening, and incorporate reflection or discussion about transcripts afterward. Such structured support reduces cognitive overload and deepens understanding of text organization and function. These strategies are consistent with metacognitive instruction as articulated by Goh & Vandergrift (2021) and supported by recent digital listening studies showing significant gains in comprehension and metacognitive awareness (Syahfutra et al., 2023; Khairat, 2024).

Interactive platforms such as ESLVideo.com can also enhance comprehension by providing instant feedback and active engagement opportunities. When integrated into collaborative tasks, they promote communicative interaction and peer learning, echoing Vygotsky's (1978) concept of the Zone of Proximal Development. However, the study also confirms that effective ICT integration requires consistent teacher facilitation and clear classroom management strategies to prevent distraction and ensure equitable participation.

Moreover, the findings highlight the necessity of incorporating digital literacy development within vocational English curricula. As seen in Cycle I, limited familiarity with ICT platforms hindered participation; by Cycle II, guided modeling and peer mentoring resolved these issues, enabling more autonomous learning. This outcome aligns with recent research showing that digital competence—both student and teacher—is foundational for meaningful technology-enhanced language learning, through enhanced implementation, autonomy, and sustainable practice (Feng & Sumettikoon, 2024).

Limitations

Despite its strengths, the study has certain limitations. The research involved only one class of 36 students from a specific vocational program, which restricts generalizability. Without a control group, it is impossible to isolate the exact magnitude of improvement directly attributable to the intervention. Additionally, the dual role of the researcher as both instructor and observer introduces the possibility of subjective interpretation bias.

These limitations, however, are typical in Classroom Action Research (CAR), which prioritizes contextual authenticity and ecological validity over experimental control (Burns, 2010). Future studies could strengthen external validity by using mixed-method or quasi-experimental designs across multiple schools or disciplines. Longitudinal research could also examine whether the observed gains persist over time or transfer to other English language skills such as speaking and reading comprehension.

Recommendations

Based on the findings, several recommendations can be made for both practice and research. Teachers are encouraged to institutionalize reflective teaching cycles as part of continuous professional development, enabling iterative refinement of ICT-based methods. Teacher training programs should balance

technological competence with pedagogical scaffolding skills to ensure that ICT integration supports—not replaces—effective teaching.

Researchers could extend this study by exploring ICT integration across a broader range of vocational domains, including hospitality, engineering, and health sciences, to test the adaptability of multimodal instructional models. Incorporating metacognitive strategy instruction within ESP curricula could also foster greater learner autonomy and comprehension monitoring.

Educational policymakers should support ICT-based language learning by ensuring digital infrastructure, equitable access, and institutional training. These measures would enable schools—particularly vocational ones—to sustain digital innovation while minimizing disparities in readiness and resources.

Conclusion

This study investigated how integrating YouTube and ESLVideo.com within a Genre-Based Approach (GBA) through Classroom Action Research (CAR) can improve vocational students' English listening comprehension. Across two cycles, the intervention showed that authentic multimodal materials, when scaffolded effectively, enhanced comprehension, motivation, and learner autonomy. The steady increase in listening scores and observed behavioral changes indicate that combining visual and auditory input reduced cognitive load and supported active meaning-making.

The findings extend current perspectives in technology-enhanced language learning, showing that successful ICT use depends on pedagogical design, not the technology itself. The GBA framework provided the structure necessary to transform authentic audiovisual content into comprehensible and purposeful learning input. This outcome supports contemporary theories of multimedia learning (Mayer, 2009), cognitive load management (Sweller, 2019), and self-determination (Deci & Ryan, 2000), emphasizing that comprehension and motivation grow when instruction is scaffolded, contextualized, and learner-centered.

Pedagogically, the study highlights the value of reflective, context-driven design in vocational English teaching. Integrating field-relevant materials and interactive tasks connected classroom learning with real workplace communication, while iterative reflection helped refine strategies to overcome technical and motivational challenges. These findings demonstrate how CAR can operationalize continuous improvement in ICT-based instruction.

Nonetheless, the study's scope—limited to one class and lacking a control group—constrains broader generalization. The dual role of the researcher as instructor and observer may also have influenced outcomes, though triangulated data minimized bias. Future studies should employ larger samples, multi-site designs, or longitudinal approaches to examine the sustainability and transfer of learning gains. Exploring emerging digital tools such as AI-based listening support

could further enrich understanding of differentiated instruction in ESP contexts.

In conclusion, this study affirms that interactive ICT-based listening instruction, when grounded in theory and reflective pedagogy, can transform vocational classrooms into active, learner-centered environments. The model presented offers a practical framework for integrating authentic multimedia to build comprehension, confidence, and autonomy in vocational English learning.

Acknowledgement

The researcher expresses sincere gratitude to SMK N 2 Depok, Yogyakarta, for granting permission and support during the research process. Special appreciation is extended to the English teacher and students of Class XI Computer and Network Engineering, whose active participation made this study possible. Deep thanks are also conveyed to the Department of English Education, Yogyakarta State University, for its academic guidance and constructive feedback throughout the research.

References

- Adi, M., Nasrullah, N., & Rosalina, E. (2022). Investigating EFL students' listening comprehension: Problems and causes. *International Journal of Education, Language and Literature*, 2(1), 12-25. <https://doi.org/10.26740/elitejournal.v2n1.p12-25>
- Al-Jumaily, I. H., & Alazzawi, I. T. (2025). The influence of multimodal visual methodologies on EFL University students' audio-visual comprehension, verbal and nonverbal communication. *Journal of Language and Education*, 11(2), 35-56. <https://doi.org/10.17323/jle.2025.22731>
- Aziz, A. F., & Anjaniputra, A. G. (2025). The Alignment of ESP in Teaching Activities and Materials at Vocational High School. *Tamaddun*, 24(1), 60-72. <https://doi.org/10.33096/tamaddun.v24i1.918>
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Brown, H. D., & Lee, H. (2015). *Teaching by Principles: An Interactive Approach to Language Pedagogy*. Pearson.
- Burns, A. (2010). *Doing Action Research in English Language Teaching*. Routledge.
- Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227-268. https://doi.org/10.1207/S15327965PLI1104_01
- Denzin, N. K. (2017). *The Research Act: A Theoretical Introduction to Sociological Methods*. Routledge.
- Dudley-Evans, T., & St. John, M. J. (1998). *Developments in English for Specific Purposes: A Multi-Disciplinary Approach*. Cambridge University Press.
- Efendi, A. Y., & Astutik, Y. (2024). Podcasts Boost Listening Skills in Vocational High School Students: Podcasting Improves Listening Skills for Vocational High School Students. *Indonesian Journal of Education Methods Development*, 19(3). <https://doi.org/10.21070/ijemd.v19i3.845>

- Febriani, H., Asvio, N., Rahmadoni, J., & Vivekanantharasa, R. (2022). Improving EFL learners listening skills by using audio visual aids. *Linguists: Journal Of Linguistics and Language Teaching*, 8(2), 216. <https://doi.org/10.29300/ling.v8i2.7662>
- Feng, L., & Sumettikoon, P. (2024). An empirical analysis of EFL teachers' digital literacy in Chinese higher education institutions. *International Journal of Educational Technology in Higher Education*, 21(1). <https://doi.org/10.1186/s41239-024-00474-1>
- Field, J. (2008). *Listening in the Language Classroom*. Cambridge University Press.
- Gardner, R. C., & Lambert, W. E. (1972). *Attitudes and Motivation in Second Language Learning*. Newbury House Publishers.
- Goh, C. C. M., & Vandergrift, L. (2021). *Teaching and learning second language listening: Metacognition in action*. Routledge.
- Hadist, M. K., Hidjanah, H., Dtakiyatuddaaimah, D., & Dewanti, L. (2022). An analysis of EFL students' difficulties in listening comprehension at STKIP Muhammadiyah Bogor. *Jurnal Education FKIP UNMA*, 8(1), 174-179. <https://doi.org/10.31949/educatio.v8i1.1847>
- Husnawati, N., Puspitasari, M., & Yundayani, A. (2024). Enhancing intrinsic motivation and listening skills through integrating authentic learning materials. *English Review: Journal of English Education*, 12(2), 559-566. <https://doi.org/10.25134/erjee.v12i2.9761>
- Kemmis, S., & McTaggart, R. (1988). *The action research planner*. Deakin University Press.
- Kemmis, S., McTaggart, R., & Nixon, R. (2014). *The action research planner: Doing critical participatory action research*. Springer.
- Khairat, N. (2024). The use of YouTube to enhance learners' English listening skills. *Surakarta English and Literature Journal*, 7(2), 134-147. <https://doi.org/10.52429/selju.v7i2.265>
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
- Lertchalermtipakoon, P., Wongsunbun, U., & Kawinkoonlasate, P. (2021). Need analysis: English language use by students in the tourism and hospitality and industry. *English Language Teaching*, 14(3), 59. <https://doi.org/10.5539/elt.v14n3p59>
- Liu, T., Zhang, Z., & Gao, X. (. (2023). Pedagogical design in technology-enhanced language education research: A scoping review. *Sustainability*, 15(7), 6069. <https://doi.org/10.3390/su15076069>
- Masruddin, M., & Nasriandi, N. (2022). Lexical and Syntactical Errors Performed by Junior High School Student in Writing Descriptive Text. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 10(1),

1094-1100.

- Masruddin, M., Amir, F., Langaji, A., & Rusdiansyah, R. (2023). Conceptualizing linguistic politeness in light of age. *International Journal of Society, Culture & Language*, 11(3), 41-55.
- Mayer, R. E. (2009). *Multimedia learning*. Cambridge University Press.
- Nasution, Z. & Prihatini, S. (2024). Enhancing English listening skills through metacognitive strategy: A study on student achievement. *EXCELLENCE: Journal of English and English Education*, 5(1). <https://doi.org/10.47662/ejee.v5i1.1196>
- Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
- Pansri, B., Thongchotchat, V., Kurashige, K., Watanabe, S., & Sato, K. (2025). Self-regulated learning in action: Empowering physics curriculum with ICT and practical approaches. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-025-09854-6>
- Pei, T., Suwanthep, J., & Lu, H. (2023). The effect of self-directed online metacognitive listening practice on Chinese EFL learners' listening ability, metacognition, and self-efficacy. *Frontiers in Psychology*, 14. <https://doi.org/10.3389/fpsyg.2023.1285059>
- Purwanto, B. A. (2024). Enhancing listening ability by using short video among vocational school students. *LINCA: Jurnal Kajian Bahasa*, 1(2). <https://doi.org/10.18178/ijlll.2024.10.1.480>
- Santika, S., Wirza, Y., & Emilia, E. (2022). A needs analysis for ESP in a multimedia study programme of a vocational high school. *International Journal of Education*, 15(1), 61-68. <https://doi.org/10.17509/ije.v15i1.46156>
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Setiyawan, H., Suharno, S., & Pambudi, N. A. (2023). The influence of digital and vocational information literacy on student learning outcomes. *Jurnal Pendidikan Vokasi*, 13(2), 192-204. <https://doi.org/10.21831/jpv.v13i2.53999>
- Shamsi, E., & Bozorgian, H. (2024). Collaborative listening using multimedia through metacognitive instruction: A case study with less-skilled and more-skilled EFL learners. *Asian-Pacific Journal of Second and Foreign Language Education*, 9(1). <https://doi.org/10.1186/s40862-023-00248-8>
- Shao, Y., Wu, J., Li, Y., Lu, Q., & Wang, Z. (2025). The impact of digital technology use on EFL students' English academic performance: The mediating roles of emotional intelligence and learning engagement. *BMC Psychology*, 13(1). <https://doi.org/10.1186/s40359-025-02967-8>
- Siregar, S. H. (2024). The effectiveness of communicative approach in teaching present tense to vocational college students: An experimental study. *English Teaching and Linguistics Journal*, 5(2). <https://doi.org/10.30596/etlij.v5i2.23244>

- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295-312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Syahfutra, W., Safriyanti, M., & Yusandra, T. F. (2023). Exploring Vandergrift 'S Metacognitive strategies use and its impact on listening comprehension: A comprehensive review. *International Journal of Language Pedagogy*, 3(2), 43-51. <https://doi.org/10.24036/ijolp.v3i2.53>
- Valentin, O. (2024). Improving listening skill with a video based podcast for vocational high school students'. *Journal of English Teaching and Learning*, 3(2), 73-81. <https://doi.org/10.62734/jetling.v3i2.593>
- Verbeke, G., & Simon, E. (2023). Listening to accents: Comprehensibility, accentedness and intelligibility of native and non-native English speech. *Lingua*, 292, 103572. <https://doi.org/10.1016/j.lingua.2023.103572>
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
- Warschauer, M. (2011). *Learning in the Cloud: How (and Why) to Transform Schools with Digital Media*. New York: Teachers College Press.
- Zhang, Y., & MacWhinney, B. (2023). The role of novelty stimuli in second language acquisition: Evidence from the optimized training by the pinyin tutor at TalkBank. *Smart Learning Environments*, 10(3). <https://doi.org/10.1186/s40561-023-00223-3>
- Zhang, Z., & Crawford, J. (2023). EFL learners' motivation in a gamified formative assessment: The case of Quizizz. *Education and Information Technologies*, 29(5), 6217-6239. <https://doi.org/10.1007/s10639-023-12034-7>
- Zuo, M., Zhong, Q., Wang, Q., Yan, Y., Liang, L., Gao, W., & Luo, H. (2024). Supporting home-based self-regulated learning for secondary school students: An educational design study. *Sustainability*, 16(3), 1199. <https://doi.org/10.3390/su16031199>