

## Bridging the Gap: A Dual-Perspective Analysis of Audio-Visual Material Integration in Indonesian EFL Speaking Classrooms

Fikra Afdhali<sup>1</sup>, Utami Dewi<sup>2</sup>, Farida Repelitawaty BR Kembaren<sup>3</sup>

<sup>1,2,3</sup>Faculty Of Tarbiyah Science and Teachers' Training, Universitas Islam Negeri Sumatera Utara

### Article Info

#### Article history:

Received 2026-01-05

Revised 2026-01-28

Accepted 2026-02-03

#### Keywords:

Audio-Visual Materials  
EFL Speaking Skills  
Implementation Gap  
Indonesian Context  
Student Perceptions  
Teacher Perceptions  
TPACK

### ABSTRACT

The integration of English audio-visual (AV) materials has been widely promoted as an effective approach for enhancing speaking skills in English as a Foreign Language (EFL) contexts; however, a persistent gap remains between its theoretical potential and classroom implementation in Indonesia. This qualitative case study aims to investigate perceptions and challenges of using AV materials for speaking skill development, examining both student and teacher perspectives. The study employed an applied qualitative research design, collecting data through Google Forms, classroom observations, and document analysis, involving two English teachers and 13 students at a private language institute in Lhokseumawe, Indonesia. Thematic analysis revealed a systemic misalignment between stakeholders. Students perceived AV materials as engaging, anxiety-reducing, and essential for accessing authentic language input, and they expressed a strong preference for interactive and creative speaking activities. In contrast, teachers, despite recognizing the pedagogical value of AV materials, faced significant constraints, including limited preparation time, unstable technological infrastructure, exam-oriented curricular pressures, and insufficient Technological Pedagogical Content Knowledge (TPACK). This TPACK deficiency led to predominantly input-focused and pedagogically simplified use of AV materials, limiting opportunities for meaningful spoken output. The study concludes that the primary barrier to effective AV integration lies not in technological access but in interconnected institutional, professional, and curricular constraints. To address this gap, the study proposes a tripartite framework consisting of institutional support re-engineering, TPACK-oriented professional development, and curricular realignment to legitimize AV-mediated speaking practice. These findings provide context-sensitive recommendations for transforming AV use from passive consumption into active speaking production in Indonesian EFL classrooms.

*This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Fikra Afdhali

Faculty of Tarbiyah Science and Teacher's Training, Universitas Islam Negeri Sumatera Utara Medan

Email: [fikraafdjali91@gmail.com](mailto:fikraafdjali91@gmail.com)

## 1. INTRODUCTION

The development of oral proficiency remains a formidable challenge in global English as a Foreign Language (EFL) contexts, particularly in environments such as Indonesia, where authentic exposure is limited. The Indonesian educational landscape, characterized by exam-oriented curricula, large class sizes, and teacher-centered pedagogy, often marginalizes communicative speaking practice in favor of grammar and reading comprehension [1] [2]. This is compounded by affective barriers such as malu (shyness), which further inhibit students' willingness to communicate [3].

Speaking is a complex, real-time, productive skill that requires the simultaneous integration of grammatical, discourse, sociolinguistic, and strategic competences [10]. In Indonesia, development is hindered by a confluence of factors: exam washback that prioritizes non-communicative skills, large classes, teacher-centered methods, and potent affective filters such as malu [11], [12].

In response to these challenges, audio-visual (AV) materials have emerged as a promising pedagogical intervention. Grounded in cognitive theories like Paivio's Dual Coding Theory and Mayer's Cognitive Theory of Multimedia Learning, AV resources provide contextualized, comprehensible input and can lower the affective filter, making them theoretically potent tools for speaking development [4], [5], [6]. Audio-visual materials integrate auditory and visual channels, offering rich, authentic input. Their use is supported by theories of multimedia learning [13], comprehensible input [14], and sociocultural mediation [15].

In the digital era, Mobile-Assisted Language Learning (MALL) provides ubiquitous access to resources on platforms such as YouTube and TikTok, enabling innovative practices such as shadowing and dubbing [16], [17]. International and Indonesian research has documented the benefits of AV materials for vocabulary acquisition, learner motivation, and classroom engagement [7], [8].

Despite these documented advantages, a critical synthesis reveals a persistent gap between research and practice. Students generally perceive AV materials favorably, reporting increased motivation, reduced anxiety, and benefits for pronunciation and vocabulary [18], [19]. Teachers likewise acknowledge these benefits but report significant implementation barriers, including time constraints, inadequate training, infrastructural issues, and a lack of pedagogical knowledge for designing effective AV-based speaking tasks [20], [21]. Existing studies have predominantly focused on quantitative learning outcomes or single-perspective views (either teacher or student), overlooking the dynamic interplay between these key stakeholders. This omission is significant because successful technology integration hinges not only on theoretical efficacy but also on user perceptions and enacted practices [9].

The Technological Pedagogical Content Knowledge (TPACK) framework [22] is crucial for analyzing this disconnect. Effective integration requires teachers to synthetically combine knowledge of technology (TK), pedagogy for teaching speaking (PK), and English language content (CK). A deficit in integrated TPACK often leads to the use of AV as a passive presentation tool rather than a scaffold for communicative output. Consequently,

---

there is a limited understanding of the aligned or divergent perceptions teachers and students hold regarding AV use for speaking, and a lack of holistic, qualitative insights into the lived classroom realities within Indonesia's unique EFL ecosystem.

Addressing this gap is an urgent pedagogical imperative. As Indonesia undergoes a digital transition in education, without understanding user perceptions, investments in technology risk leading to 'implementation failure' where tools are underutilized or misused, perpetuating speaking deficits and teacher frustration. Therefore, this study is not merely academic but a necessary investigation to inform sustainable, human-centered technology integration. This study is guided by research questions focusing on students' perceptions of the utilization of English audio-visual materials in EFL speaking classes, teachers' perceptions of the use of audio-visual materials in EFL speaking skill development, and the challenges faced by both teachers and students when using audio-visual materials in EFL speaking classes.

## **2. METHOD**

### **2.1 Research Design**

This study employed a qualitative case study design with an exploratory approach, conducted at Eka Learning Center (ELC), a private language institute in Lhokseumawe, Aceh.

### **2.2 Participants**

Participants were selected via purposive sampling:

- *Teachers*: Two female English teachers with over three years of experience.
- *Students*: Thirteen students (ages 11-20) with proficiency levels ranging from elementary to upper-intermediate, drawn from various educational levels.

### **2.3 Data Collection**

Triangulation was achieved through three methods:

1. *Google Form Questionnaires*: Separate open-ended questionnaires for teachers and students to explore perceptions, benefits, challenges, and preferences.
2. *Classroom Observations*: Four non-participant observations over one month, using a structured checklist to document AV implementation, student engagement, and teacher strategies.
3. *Document Analysis*: Analysis of teaching modules (*Rencana Pelaksanaan Pembelajaran/RPP*) to examine the planned integration of AV materials and task design.

### **2.4 Data Analysis**

Data were analyzed using thematic analysis following Braun and Clarke's six-phase framework [23]. Data from questionnaires, observation notes, and documents were coded, categorized, and reviewed to identify recurring themes related to perceptions and challenges. Triangulation across data sources and member checking were employed to ensure credibility and trustworthiness [24]

---

### 3. RESULTS AND DISCUSSION

#### 3.1. Results

##### 3.1.1 Students' Perceptions: Affective Engagement and a Demand for Agency

Students reported overwhelmingly positive perceptions, centered on affective and cognitive benefits.

- *Affective Benefits*: Students described AV-based lessons as more interesting and less boring, leading to reduced speaking anxiety and increased confidence. Observations confirmed more voluntary participation during AV activities.
- *Cognitive Benefits*: Students demonstrated metacognitive awareness, valuing AV materials for exposure to authentic native-speaker pronunciation, everyday vocabulary, and pragmatic language use. They reported learning through imitation and shadowing.
- *Preferences and Expectations*: Students showed a strong preference for contemporary, short-form content from platforms like YouTube and TikTok. Crucially, they expressed a desire to move beyond passive viewing, suggesting active tasks like role-plays, simulations, and creating their own videos.
- *Challenges*: Students identified fast speech, difficult vocabulary, and an unstable internet connection as the primary obstacles.

##### 3.1.2 Teachers' Perceptions: A Theory-Practice Divide

A pronounced gap between theoretical endorsement and practical constraint characterized teachers' perceptions.

- *Theoretical Endorsement*: Teachers strongly affirmed the value of AV materials for providing authentic input, cultural context, and increasing student engagement.
- *Constrained Implementation*: Despite this belief, their practice was mediated by significant barriers:
  1. *Time Burden*: The exhaustive process of curating, previewing, and aligning AV materials was cited as the primary deterrent.
  2. *TPACK Deficit*: Teachers lacked integrated knowledge to design scaffolded, output-oriented speaking tasks based on AV input. Document analysis of RPPs revealed activities often stopped at passive comprehension checks (e.g., "answer questions") rather than progressing to creative speaking production.
  3. *Systemic Barriers*: Unreliable infrastructure (internet, equipment), heterogeneous student proficiency levels, and the washback effect of exam-oriented curricula further discouraged consistent and innovative use.

##### 3.1.3 The Systemic Perception Gap and Its Manifestation

The core finding was a significant misalignment between student and teacher perspectives, constituting a systemic implementation gap.

---

- Students, as digital natives, viewed AV as a central, engaging platform for authentic learning and expected dynamic, interactive use.

Teachers, burdened by logistical constraints, pedagogical uncertainty, and systemic pressures, adopted a cautious, supplemental, and often passive approach to AV integration. This gap ensured AV's potential as a speaking catalyst remained largely untapped. Observations confirmed that while AV boosted engagement, post-viewing activities often lacked the structured scaffolding necessary to transform input into meaningful spoken output.

### 3.2. Discussion

The findings illuminate the complex ecosystem constraining AV integration in Indonesian EFL speaking classes. Students' positive perceptions validate cognitive and affective theories, demonstrating that AV content is relevant and lowers anxiety while providing dual-coded input. Through visual context and authentic language exposure, students can process meaning more efficiently, reducing cognitive overload, which is particularly beneficial in EFL contexts with limited real-life exposure. Their desire for active production aligns with Swain's Output Hypothesis [25], suggesting that learners do not merely want to consume multimedia input passively but expect opportunities to transform that input into spoken output through interaction, rehearsal, and feedback. This indicates that AV materials are most pedagogically effective when positioned as catalysts for communicative tasks rather than endpoints of instruction.

However, the theory-practice divide experienced by teachers reveals the insufficiency of theoretical potential alone. While teachers conceptually acknowledge the value of AV materials, their classroom enactment is mediated by structural, institutional, and personal constraints. The constraints they face, particularly time poverty and a TPACK deficit, represent classic first- and second-order barriers to technology integration [26]. First-order barriers, such as limited instructional time, rigid syllabi, and curriculum demands, restrict experimentation with AV-based speaking tasks. Second-order barriers, including beliefs about teaching effectiveness and limited confidence in pedagogical technology use, further inhibit meaningful integration.

The TPACK gap is critical: possessing TK, PK, and CK in isolation is insufficient without the integrative knowledge to design instructional sequences that deliberately bridge AV input with spoken output. Teachers may be able to operate digital tools (TK), understand communicative language teaching principles (PK), and master English content (CK), yet still struggle to orchestrate these domains into coherent speaking activities. As a result, AV materials are often used reductively, functioning primarily as listening comprehension aids or visual explanations rather than as scaffolds for interaction, meaning negotiation, or oral production. This leads to pedagogical simplification, where AV becomes an input source rather than a conversation springboard.

The washback effect from high-stakes examinations further institutionalizes this divide, subtly prioritizing grammar accuracy and reading comprehension over communicative competence. Within such an assessment-driven environment, teachers are incentivized to allocate classroom time toward test-relevant skills, even when they recognize

the long-term importance of speaking proficiency. Consequently, AV-based speaking activities are perceived as supplementary or expendable rather than integral, reinforcing conservative instructional practices.

This dynamic creates a vicious cycle: contextual and institutional constraints limit opportunities for communicative practice, which in turn inhibits the development of teachers' integrated TPACK. Limited pedagogical success with AV results in suboptimal learning outcomes, reinforcing the perception that AV materials are inefficient or non-essential for speaking instruction. Over time, this cycle normalizes minimal AV use and perpetuates speaking deficits among learners, despite positive student attitudes and strong theoretical justification. Breaking this cycle requires not only access to technology but also sustained professional development and systemic alignment between curriculum goals, assessment practices, and pedagogical innovation.

#### **4. CONCLUSION**

This study underscores that the challenge of integrating audio-visual (AV) materials for speaking development in Indonesian EFL contexts is fundamentally systemic rather than technological. The central conclusion is that a misalignment exists between students' readiness to engage in digitally mediated speaking practices and teachers' constrained capacity to facilitate such learning effectively. While learners are prepared to participate actively and creatively, instructional practices remain limited by structural and professional conditions that inhibit pedagogically meaningful integration.

From both theoretical and practical standpoints, this research has several important implications. First, it reinforces the relevance of the Technological Pedagogical Content Knowledge (TPACK) framework by demonstrating how deficiencies in integrated knowledge translate into instructional decisions that prioritize exposure over production. Second, it suggests that professional development initiatives should move beyond technical training toward sustained, practice-oriented support that enables teachers to design AV-mediated speaking tasks aligned with communicative objectives. Third, at the institutional level, the findings imply the need for policy and curriculum structures that legitimize speaking-oriented, technology-enhanced pedagogy rather than privileging exam-driven outcomes alone.

Several limitations bound this study. It was conducted within a single private language institute with a small number of participants, which restricts the generalizability of the findings across diverse educational settings. Additionally, the focus on perceptions and classroom practices did not extend to longitudinal measurement of speaking proficiency gains, limiting claims regarding long-term learning outcomes.

Future research is encouraged to expand this inquiry across multiple institutions and educational levels, including public schools and higher education contexts, to explore whether similar systemic patterns persist. Mixed-method or longitudinal studies could further examine how targeted TPACK-based interventions influence teachers' instructional practices and students' speaking development over time. For the general public and educational stakeholders, this research contributes to a clearer understanding that successful educational technology integration depends not merely on access to digital tools, but on

---

supportive ecosystems that empower teachers and meaningfully engage learners. In doing so, it advocates for a shift from passive media consumption toward active, communicative language use that better prepares learners for real-world English communication.

## REFERENCES

- [1] Mustafa, F., & Yusuf, Y. Q. (2022). The washback effect of high-stakes tests on teachers' instructional practices: The case of the Indonesian national exam. *Studies in Educational Evaluation*, 74, 101181. <https://doi.org/10.1016/j.stueduc.2022.101181>.
  - [2] Zein, S. (2017). Elementary ENGLISH education in Indonesia: Policy developments, current practices, and future prospects. *ENGLISH Today*, 33(1), 53–59. <https://doi.org/10.1017/S0266078416000417>.
  - [3] Rachmawaty, N., & Ariani, S. (2023). Willingness to communicate in Indonesian EFL classrooms: The interplay of motivation, anxiety, and self-perceived communicative competence. *Journal on ENGLISH as a Foreign Language*, 13(1), 1-23. <https://doi.org/10.23971/jefl.v13i1.4567>.
  - [4] Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
  - [5] Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>.
  - [6] Krashen, S. D. (1985). *The input Hypothesis: Issues and implications*. Longman.
  - [7] Almurashi, W. A. (2016). The effective use of YouTube videos for teaching ENGLISH language in classrooms as supplementary material. *International Journal of ENGLISH Language and Linguistics Research*, 4(3), 32–47.
  - [8] Xie, Y., Chen, F., & DeMarinis, M. (2022). Mobile language learning applications: A meta-analysis. *Frontiers in Psychology*, 13, 947975. <https://doi.org/10.3389/fpsyg.2022.947975>
  - [9] Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education*, 42(3), 255-284. <https://doi.org/10.1080/15391523.2010.10782551>
  - [10] Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47. <https://doi.org/10.1093/applin/I.1.1>
  - [11] Bygate, M. (2019). *Speaking*. Routledge. <https://doi.org/10.4324/9781315833546>
  - [12] Lestari, S. (2019). The washback effect of national examination on EFL teaching in Indonesian secondary schools. *Journal of ENGLISH Education and Teaching*, 3(1), 1-18. <https://doi.org/10.33369/jeet.3.1.1-18>
  - [13] Zein, S. (2017). Elementary ENGLISH education in Indonesia: Policy developments, current practices, and future prospects. *ENGLISH Today*, 33(1), 53–59. <https://doi.org/10.1017/S0266078416000417>
  - [14] Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
  - [15] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
  - [16] Garzón, J., Soria, J., & Baldiris, S. (2023). Video-based learning in second language acquisition: A systematic review. *Computer Assisted Language Learning*, 36(5-6), 510–534. <https://doi.org/10.1080/09588221.2021.1888757>
  - [17] Kukulka-Hulme, A., & Viberg, O. (2023). Mobile-Assisted language learning: Current trends and future directions. *ReCALL*, 35(1), 5–24. <https://doi.org/10.1017/S0958344022000214>
  - [18] Zhou, M., Wu, J., & Li, H. (2021). Video-based language learning and speaking proficiency: A mixed-method study. *Computer Assisted Language Learning*, 34(6), 733–755. <https://doi.org/10.1080/09588221.2019.1688830>
  - [19] Rahman, M., & Karim, A. (2023). Reducing speaking anxiety through video-based tasks in EFL classrooms. *Innovation in Language Learning and Teaching*, 17(4), 357–371. <https://doi.org/10.1080/17501229.2022.2089017>
  - [20] Wulandari, N., & Sari, D. (2022). Teachers' challenges in integrating technology for EFL speaking practice in Indonesia. *Journal of Language and Education*, 8(3), 45–56. <https://doi.org/10.17323/jle.2022.15256>
  - [21] Garzón, E., Saz, A., & Rodríguez, J. M. (2023). Teacher and learner perceptions of mobile-Assisted
-

- language learning in speaking practice. *Computer Assisted Language Learning*, 36(7), 1338–1357. <https://doi.org/10.1080/09588221.2022.2039823>
- [22] Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- [23] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [24] Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.
- [25] Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *input in second language acquisition* (pp. 235–253). Newbury House.
- [26] Ertmer, P. A. (1999). Addressing first- and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development*, 47(4), 47–61. <https://doi.org/10.1007/BF02299597>
-