



# The Role of Audiovisual Materials in Enhancing English Speaking Skills: A Systematic Review (2020–2026)

Anwer Mohammed Elhassan Hussein

Research Scholar, Faculty of Graduate Studies, Alzaiem Alazhari University, Sudan  
Email: [anwermohamad79@gmail.com](mailto:anwermohamad79@gmail.com)

Received: 25 Aug 2026, Received in revised form: 21 Sep 2026, Accepted: 28 Sep 2026, Available online: 04 Oct 2026

©2026 The Author(s). Published by IJTLE. This is an open-access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>).

## Abstract

*Speaking is among the hardest skills for English learners to develop, especially where authentic oral interaction is scarce. Audiovisual (AV) materials such as films, YouTube videos, TED Talks and multimedia courseware are increasingly used to expose learners to natural pronunciation, intonation and discourse. This review follows PRISMA 2020 reporting principles to synthesize peer-reviewed studies from 2020–2026 on AV materials and English speaking outcomes. Fourteen documents were assessed against pre-specified criteria, and four studies (Indonesia, Bangladesh; n = 6 to 50) were included. Every study favored AV-supported instruction for fluency, pronunciation, confidence, interaction and overall oral performance, but the evidence is weak: two studies reported scores, none had a comparison group, baseline or follow-up, and the remaining evidence is self-reported. No study measured grammatical accuracy or tested a moderator. The findings are consistent with Dual Coding Theory, Mayer's multimedia learning theory, and the Input and Output Hypotheses, yet cannot distinguish among them. The review identifies methodological gaps and recommends controlled studies that vary one component at a time (task, audio-only versus AV, captioning) with reliable speaking ratings and delayed post-tests.*

**Keywords—** audiovisual materials, English speaking skills, oral proficiency, EFL/ESL learners, multimedia learning, systematic review, PRISMA

## I. INTRODUCTION

### 1.1 Background and Context

Speaking is widely regarded as the most demanding skill for English as a Foreign Language (EFL) and Second Language (ESL) learners, because oral production requires real-time coordination of vocabulary, grammar, pronunciation and pragmatic competence under social pressure. Limited exposure to natural spoken English, low confidence and high anxiety contribute to persistently weak speaking outcomes (Kathirvel & Hashim, 2020). AV materials are positioned as a technology-based strategy because learners can see and hear pronunciation, stress and intonation

simultaneously and then imitate them, which may reduce anxiety and build confidence. Practical constraints recur in the literature: difficulty sourcing suitable materials, insufficient teacher training, and unequal access to equipment and connectivity (Kathirvel & Hashim, 2020).

Empirical work since 2020 spans many contexts. In a quasi-experiment with 80 Jordanian university students, a YouTube-based group outperformed a traditional-instruction group, with pronunciation and fluency/coherence improving most and grammatical accuracy least (Saed et al., 2021). Similar designs reported gains from authentic video among Iraqi

undergraduates (Saber, 2025) and from animation video among Indonesian tenth-graders (Fitri, 2022;  $\eta^2 = .179$ ). Task-based action research combined AV media with communicative tasks and raised class speaking means from 45.4 to 75.8 (Pinto & Lomi, 2026). Qualitative and mixed-methods studies add teacher perspectives in Namibian primary classrooms (Shikongo, 2025), multimodal gains among 214 Iraqi students (Al-jumaily & Alazzawi, 2025), and combined listening-and-speaking gains in Bangladesh (Rahman & Jamila, 2024). These sources point in a common direction but differ widely in design and outcome measure.

## 1.2 Theoretical Grounding

Five frameworks guide the review. **Dual Coding Theory** (Paivio, 1971, 1986; Clark & Paivio, 1991) holds that verbal and imagery information are processed in separate but linked systems, so a voice paired with visible articulation and context offers two routes for encoding pronunciation. **Mayer's Cognitive Theory of Multimedia Learning** (2020) adds that limited-capacity channels benefit from coherent, well-segmented input with low extraneous load, so clip length and pacing matter. **Krashen's Input Hypothesis** (1985) treats comprehensible input as the driver of acquisition, with visual context making input understandable, although speaking is seen as an eventual output. **Swain's Comprehensible Output Hypothesis** (1985) argues that learners also need to produce language to notice gaps, which supports using AV material as a platform for production tasks. **Vygotsky's Sociocultural Theory** (1978) frames AV materials as cultural tools whose value depends on teacher-mediated, socially structured activity. The first two explain cognitive mechanisms, the next two offer partly competing accounts of exposure versus production, and the last situates both in social interaction.

## 1.3 Problem, Purpose and Research Questions

Existing evidence is dispersed, drawn from varied ages and contexts, and operationalizes "speaking improvement" in incompatible ways: IELTS band descriptors, effect sizes, action-research score bands, teacher perception and composite communication scores. Some studies bundle speaking with listening. A transparent, criteria-driven review is needed to determine which speaking dimensions are most reliably enhanced, under what conditions, and for which learners. The purpose is to identify, appraise and synthesize peer-reviewed empirical evidence from

2020–2026 on how AV materials affect English speaking among EFL/ESL learners. Four questions guide the review:

- **RQ1:** What types of AV materials and approaches (passive viewing versus active production) have been used?
- **RQ2:** Which speaking dimensions (fluency, pronunciation, accuracy, confidence, interaction) are most positively affected?
- **RQ3:** Which conditions (task integration, teacher mediation, feedback, age/proficiency, resources) moderate effectiveness?
- **RQ4:** What methodological trends and gaps characterize the research, including whether speaking can be isolated from other skills?

The review is delimited to studies published 2020–2026 that report a measured or assessed speaking-related outcome in EFL/ESL learners of any age. It aims to give curriculum designers and teachers evidence-based guidance and to consolidate a fragmented post-pandemic literature.

## II. LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

### 2.1 Defining Audiovisual Materials

AV materials are tools that combine sound and image to make language input more concrete, authentic and memorable than text alone (Kathirvel & Hashim, 2020). For speaking, two functions are useful to distinguish. As **modeling input**, learners watch and absorb pronunciation, intonation and discourse patterns through exposure (Saed et al., 2021; Saber, 2025). As a **platform embedded in production tasks**, the material is woven into cycles of communicative activity or used as a live scaffold for imitation and role-play (Pinto & Lomi, 2026; Shikongo, 2025). The distinction maps onto the Input-Output contrast: passive-viewing designs align with an Input account, task-integrated designs with a Comprehensible Output account, and teacher-mediated use with Sociocultural Theory.

### 2.2 Speaking Sub-constructs

Studies rarely treat speaking as a single outcome, and results are not uniform. **Fluency** appears in nearly every study and improved markedly in YouTube-based and task-based interventions (Saed et al., 2021; Pinto & Lomi, 2026). **Pronunciation** is the construct most consistently linked to AV input, consistent with the theoretical expectation that synchronized audio-visual

modeling supports encoding of segmental and prosodic features. **Grammatical and lexical accuracy** appears less responsive: gains were present but smaller than for fluency and pronunciation in Saed et al. (2021), a pattern resting on a single study. **Communicative confidence** is both an outcome and a mediator; hearing correct pronunciation in advance may indirectly build confidence and reduce shyness (Kathirvel & Hashim, 2020; Pinto & Lomi, 2026), though teacher-perceived confidence (Shikongo, 2025) is not a validated measure. **Interactional competence** is rarely isolated; Al-jumaily and Alazzawi (2025) combined verbal and nonverbal communication, making it only a provisional data point.

### 2.3 Instructional Approaches and Prior Reviews

Four patterns appear: video as passive modeling input replacing traditional content (Saed et al., 2021; Saber, 2025; Fitri, 2022); AV embedded in task cycles (Pinto & Lomi, 2026); teacher-mediated modeling for young learners through rhymes and role-play (Shikongo, 2025); and multimodal, captioned input with bundled outcomes (Al-jumaily & Alazzawi, 2025). “AV materials for speaking” is therefore a family of approaches whose design may moderate effects. Prior systematic work has concentrated on listening: Shaojie, Samad and Ismail (2022) reviewed 46 studies on AV input and listening comprehension and, by design, excluded speaking.

Speaking-focused reviews exist but are limited to YouTube and appear to have used a single database (Maulidiah et al., 2026). The present review therefore claims a narrower contribution: speaking outcomes across AV media types, with explicit eligibility rules and a quality appraisal step.

## III. METHODOLOGY

### 3.1 Design, Protocol and Reporting

This is a systematic review reported with reference to PRISMA 2020 (Page et al., 2021) and PRISMA-S (Rethlefsen et al., 2021). Narrative synthesis with comparison tables was chosen, because scoping showed speaking measured in fundamentally different ways (rated band descriptors, eta-squared, class means, perceptions) and designs from quasi-experiments to qualitative case studies. Meta-analysis was planned only if at least five studies reported comparable between-group effects. The protocol is to be registered on the Open Science Framework [registration details to be added]; PROSPERO was not used because it is limited to health-related outcomes.

### 3.2 Eligibility Criteria

Criteria were defined with the PICOS framework (Table 1) and fixed before screening.

Table 1. Eligibility criteria

| Element      | Inclusion                                                                                                                 | Exclusion                                                                      |
|--------------|---------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| Population   | EFL/ESL learners of any age, level or proficiency                                                                         | Other languages; native speakers; teacher-only studies                         |
| Intervention | AV material used to develop speaking: video, film, YouTube, animation, TED Talks, courseware, captioned video, AV aids    | Audio-only (except as comparator); AV used only for other skills               |
| Comparator   | Conventional instruction, no-video, audio-only or pre/post baseline                                                       | None required; type coded                                                      |
| Outcomes     | At least one measured or assessed speaking outcome: fluency, pronunciation, accuracy, oral score, confidence, interaction | General attitudes only, with no speaking outcome                               |
| Design       | Experimental, quasi-experimental, action research, mixed-methods, qualitative                                             | Conceptual or narrative papers; systematic reviews (used for citation chasing) |
| Publication  | Peer-reviewed journal articles, Jan 2020 onward, full text, English                                                       | Theses, conference abstracts, preprints, books, duplicates                     |

Boundary rules were set in advance. Theses were excluded from synthesis because they lack uniform peer review. Narrative papers were excluded but used for citation chasing. Studies with bundled outcomes (e.g.,

listening plus speaking) were included as partial evidence only when a speaking-specific result was separately reported. Perception-only studies were eligible if peer-reviewed and in range, but were coded as

perception-based and never merged with performance data. Exclusion reasons were assigned in a fixed order: date, document type, outcome.

3.3 Sources, Search and Selection

The intended strategy searches Scopus, Web of Science, ERIC and a fourth education database, with Google Scholar as a supplement and forward and backward citation chasing. Three concept blocks are combined with AND: AV media (audiovisual, video, YouTube, film, animation, multimedia, TED talk, subtitles, captions), speaking (speaking, oral proficiency, fluency, pronunciation, spoken English) and learners (EFL, ESL, English language learners), with filters for 2020–2026, journal articles and English. Full strings and a pilot against known studies will be reported in an appendix. In the present analysis, however, the evidence base is a **supplied set of 14 documents** and not the output of a database search; no PRISMA flow diagram can therefore be reported. Screening was done by one reviewer, with dual screening and Cohen’s kappa planned for the full search.

3.4 Extraction, Appraisal and Synthesis

A structured form captured participants, design, AV type and function (modeling versus task-embedded), teacher

mediation, feedback, instruments, evidence type (performance versus perception), results and implementation constraints. Quality was appraised with the Mixed Methods Appraisal Tool 2018 (Hong et al., 2018), reporting criterion-level ratings and not an overall score; no study was excluded on quality. Synthesis followed SWiM guidance (Campbell et al., 2020), grouping by speaking sub-construct and AV function, with vote counting based on direction of effect only. Where possible, eta-squared from two-group comparisons would be converted to Hedges’ g. Sensitivity analyses excluded (a) partial or bundled studies, (b) studies with two or more “No” appraisal ratings, and (c) perception-based evidence.

IV. RESULTS

4.1 Study Selection and Characteristics

Fourteen full-text documents were assessed; four were included and ten excluded (Table 2). Seven were excluded for predating 2020, two for lacking primary data, and one for reporting listening only.

Table 2. Full-text eligibility decisions (n = 14)

| Document                                                                                                                     | Decision                       | Primary reason                                                           |
|------------------------------------------------------------------------------------------------------------------------------|--------------------------------|--------------------------------------------------------------------------|
| Irmawati, Ihsan, & Rasmi (2020)                                                                                              | Included                       | Speaking scores (comprehension, fluency) and observation                 |
| Naz (2023)                                                                                                                   | Included (partial, borderline) | Perceived interaction and participation; no speaking-specific measure    |
| Rahman & Jamila (2024)                                                                                                       | Included (partial)             | Listening and speaking assessed together; speaking reported separately   |
| Dewayanti, Artini, & Kusuma (2026)                                                                                           | Included                       | Perceived speaking, pronunciation and confidence gains from films        |
| Kathirvel & Hashim (2020); Maulidiah et al. (2026)                                                                           | Excluded                       | No primary data (narrative paper; systematic review)                     |
| Syah & Fediyanto (2026)                                                                                                      | Excluded                       | Listening outcome only                                                   |
| Mizab (2015); Bahrani & Sim (2012); Bajrami & Ismaili (2016); Macwan (2015); Kim (2015); Sabiq (2017); Polat & Erişti (2019) | Excluded                       | Published before 2020 (several also thesis, conceptual, or non-speaking) |

Of the included studies, one was published in each of 2020, 2023, 2024 and 2026. Two were from Indonesia and two from Bangladesh; two involved secondary

students and two university students; none involved primary learners. Samples ranged from 6 to 50. Table 3 summarizes their features.

Table 3. Characteristics of included studies ( $n = 4$ )

| Feature     | Irmawati et al. (2020)                | Naz (2023)                             | Rahman & Jamila (2024)                   | Dewayanti et al. (2026)       |
|-------------|---------------------------------------|----------------------------------------|------------------------------------------|-------------------------------|
| Context     | Indonesia; Grade VIII                 | Bangladesh; first-year undergraduates  | Bangladesh; first-year, pre-intermediate | Indonesia; Grade 10           |
| Sample      | 28 (one class)                        | 50 stated; 9 in focus group            | 46 (14–36 per test); 24 surveyed         | 6                             |
| Design      | Action research, 2 cycles, no control | Survey plus focus group                | Single group, 10 tests over ~3 months    | Qualitative interviews        |
| AV material | YouTube dialogues, slides             | Slides, short films, clips, Mentimeter | TED Talks, BBC Learning English clips    | Feature films (self-selected) |
| Evidence    | Performance + perception              | Perception                             | Performance + perception                 | Perception                    |

#### 4.2 Types of AV Material and Approach (RQ1)

Four different AV types appeared, one per study. Three studies combined AV material with production tasks (pair role-play, student presentations, video presentations) and teacher mediation, so effects cannot be attributed to AV alone. Only Dewayanti et al. (2026) described largely self-directed viewing, and its evidence is perception-based.

#### 4.3 Effects by Speaking Sub-construct (RQ2)

Two studies reported scores, but scales and tasks differ and results were not pooled. Irmawati et al. (2020)

reported class means rising from 57.78 to 72.67 between cycles (fluency 57.22 to 70.00); the average of the two Cycle II aspect scores is 71.67, so the gain may be 13.89, not 14.89. Rahman and Jamila (2024) reported mean test scores rising from 13.57 to 15.16 out of 20 across ten tests (+1.59 points), with dips at Tests 3, 6 and 7. Different numbers of students sat each test, tasks became harder over time, scoring criteria were not described, and neither study reported inferential statistics, so no effect sizes could be computed.

Table 4. Summary of evidence by sub-construct

| Sub-construct            | Studies | Evidence                                                                               | Strength           |
|--------------------------|---------|----------------------------------------------------------------------------------------|--------------------|
| Fluency                  | 2       | Irmawati: 57.22 to 70.00; Dewayanti: imitation of actors                               | Very limited       |
| Pronunciation            | 2       | Perception only (Irmawati interviews; Dewayanti)                                       | Very limited       |
| Accuracy                 | 0       | None                                                                                   | No evidence        |
| Confidence               | 4       | Self-reported in all four; no validated instrument                                     | Limited, favorable |
| Interaction              | 2       | Naz (about 84% disagreed multimedia is unrelated to interaction); Irmawati observation | Very limited       |
| Overall oral performance | 2       | Score increases in Irmawati and Rahman & Jamila; no comparison group                   | Very limited       |

All reported findings favored AV-supported instruction, and no study reported a null or negative result. With four studies, two perception-only and none with a comparison group, this pattern cannot be tested for publication bias and should not be read as strong

evidence of effectiveness. Learners also reported difficulty with fast dialogue, accents, idioms and slang, coping through subtitles and replaying scenes (Dewayanti et al., 2026), and initial unfamiliarity with AV-supported testing (Rahman & Jamila, 2024).

#### 4.4 Moderating Factors (RQ3)

No moderator was examined. Task design, age, proficiency, exposure, feedback, teacher mediation and captioning were described but never compared within a study. Duration ranged from about four meetings to 24 sessions over three months. Between-study differences in context, design and outcome cannot be treated as moderator effects.

#### 4.5 Methodological Patterns (RQ4)

Speaking was scored in 2 of 4 studies, and perception data appeared in all four. None reported rater reliability, a scoring rubric, a control group, a baseline, a follow-up or a validated confidence instrument; only Irmawati

isolated speaking cleanly, and only one questionnaire reported reliability ( $\alpha = .691$ ). Reporting inconsistencies were found in every study: Irmawati's Cycle II mean does not match its aspect scores; Naz's stated N of 50 conflicts with figure percentages consistent with about 39 responses; Rahman and Jamila describe "significant" gains using only descriptive statistics; and Dewayanti's tables attribute the speaking theme inconsistently.

#### 4.6 Quality Appraisal and Sensitivity Analysis

Provisional MMAT ratings by one reviewer are shown in Table 5 (Y = yes, N = no, CT = can't tell). No study met all five criteria, and two had two "No" ratings.

Table 5. Provisional MMAT (2018) ratings (single rater)

| Study (category)              | C1 | C2 | C3 | C4 | C5 | Note                                                  |
|-------------------------------|----|----|----|----|----|-------------------------------------------------------|
| Irmawati (mixed)              | CT | CT | CT | N  | N  | No baseline or comparison; concurrent changes         |
| Naz (mixed)                   | Y  | CT | CT | CT | N  | N inconsistent with figures; AV not an intervention   |
| Rahman & Jamila (descriptive) | CT | CT | CT | N  | N  | Different examinees per test; no inferential analysis |
| Dewayanti (qualitative)       | Y  | Y  | CT | CT | Y  | Six participants; few quotations                      |

Excluding partial studies (Naz; Rahman & Jamila) left Irmawati and Dewayanti, which still favored AV on fluency, pronunciation and confidence, but interaction lost its only dedicated evidence. Excluding the two studies with two "No" ratings left only perception-based evidence. Excluding perception-based evidence left two studies with score increases but no comparison group. The direction never changed, but only because all four studies favored AV.

#### 4.7 Summary Against the Research Questions

- **RQ1:** Four AV types, one each; three paired with tasks and teacher mediation.
- **RQ2:** All findings favored AV for fluency, pronunciation, confidence, interaction and overall oral performance, but evidence is limited and accuracy was not measured.
- **RQ3:** No moderator was tested.
- **RQ4:** No control group, baseline, follow-up, rater reliability or validated instrument; only two of 14 documents combined a speaking outcome, an in-range date and a performance score.

## V. DISCUSSION

### 5.1 Summary and What the Evidence Supports

Four studies favored AV-supported speaking instruction across fluency, pronunciation, confidence and interaction. The evidence is consistent with the possibility that AV-supported instruction is associated with better performance and more positive attitudes. It does not show that AV material causes gains, that it outperforms other instruction, how large any benefit is, or which design features matter. Four studies are a small base, and uniformly positive results may reflect overrepresentation of favorable work.

### 5.2 Theoretical Interpretation

**Input versus output.** The expectation that task-embedded AV should outperform passive viewing could not be tested, because no study compared the two. The two score-reporting studies used task-embedded designs and both improved, which is compatible with an output account but equally with repeated practice of the same test, teacher feedback and growing familiarity. Support from other documents is conceptual (Bajrami & Ismaili, 2016; Macwan, 2015) or secondhand (Maulidiah et al., 2026, whose seven-document base includes three reviews and which excluded studies partly for lack of effectiveness).

**Dual Coding and CTML.** Synchronized sound and image plausibly give two channels for encoding pronunciation. The data are consistent with this but do not test it: the two sources on pronunciation rely on learner reports, no study compared audio-visual with audio-only input, and the claim that AV media were “almost twice” as effective as single-channel media in Irmawati et al. has no comparison condition behind it. Imitation of actors resembles shadowing, so repetition and motivation are a rival explanation. CTML’s load concerns are echoed by difficulty with fast speech and by Irmawati’s limit of three plays, though neither was tested. A single student’s report cannot “prove” a cognitive theory, as Naz (2023) suggests.

**Proficiency and mediation.** Older documents hint that input difficulty matters: low-level learners did not improve with news but did with cartoons and films (Bahrani & Sim, 2012), and listening gains were larger at higher levels (Kim, 2015; Polat & Erişti, 2019). These concern other skills and only motivate hypotheses for speaking. Teacher mediation described in Irmawati et al. and Rahman and Jamila fits a sociocultural framing, but again formed part of a package. Overall, all five theories predict the same pattern, so none is favored; distinguishing them requires designs that vary one component at a time.

### 5.3 Comparison with Earlier Evidence

Listening reviews (Shaojie et al., 2022) found that subtitles, video type and sound-image relation shape difficulty; among included speaking studies only subtitle use appeared, as a coping strategy. Supplied listening studies had stronger designs (randomized groups, validated anxiety scales) than any included speaking study, supporting the premise that the speaking base is less developed. Maulidiah et al. (2026) concluded that YouTube is effective, especially with structured tasks, but their single-source search and exclusion of studies for “lack of effectiveness” may favor positive findings. Perception studies are consistently favorable across years: 77.5% of students saw AV aids as important for speaking (Mizab, 2015) and 83.4% rated AV materials effective for listening and speaking (Rahman & Jamila, 2024), alongside a persistent gap between endorsement and classroom provision. Attitudes, however, are not evidence of learning. Action-research designs also tend to show cycle-to-cycle gains whatever the skill (Sabiq, 2017), limiting what Irmawati’s gain can say about AV specifically.

### 5.4 Pedagogical Implications

These suggestions are reasonable practice, not effect-size findings.

- **Pair AV input with production tasks**, such as role-play from a dialogue video or a video presentation.
- **Plan teacher mediation**: limit plays, simplify instructions, use gesture instead of translation, and grade tasks from easy to complex.
- **Select materials against explicit criteria**: learner level, speech rate, specialized vocabulary, accent variety, and subtitle availability.
- **Keep the audio in view**, using shadowing or imitation of short segments rather than attending only to visuals (Kathirvel & Hashim, 2020).
- **Support offline access** through pre-downloaded clips, and **align assessment** so that speaking is regularly assessed and AV-based practice keeps its priority.

### 5.5 Gaps in the Literature

- **Longitudinal evidence**: no delayed post-test; the longest study ran about three months.
- **Instruments**: no validated confidence or anxiety scale, no reported rater reliability or rubric.
- **Populations**: no primary-school learners; only Indonesia and Bangladesh represented.
- **Accuracy and higher-order skills**: accuracy, interactional competence and discourse organization were unmeasured; the claim that grammar benefits less than fluency rests only on a secondhand summary of Saed et al. (2021).
- **Controlled comparisons**: no AV versus audio-only, passive versus task-embedded, or captioned versus uncaptioned contrasts.

### 5.6 Limitations

The 14 documents were supplied rather than retrieved through a database search, so selection was not systematic and the PRISMA flow cannot be reported. Only English full texts were considered, theses were excluded by rule (removing relevant data such as Shikongo, 2025), and seven documents were excluded for predating 2020. All four included studies, and all documents in Maulidiah et al. (2026), favored AV, so publication bias cannot be excluded. Samples were small and heterogeneous, pooling was impossible, and findings are descriptive. Screening, extraction and MMAT appraisal were done by one reviewer, and the

borderline inclusion of Naz (2023) could be reversed (leaving three studies without changing the direction of findings). Some included articles contain internal inconsistencies, and the peer-review status of some journals remains to be confirmed.

## 5.7 Conclusion

The literature tells a consistent story about attitudes and a cautiously positive story about speaking performance, but it is thin on evidence that AV material, as distinct from good teaching, is responsible. The most useful next step is a small set of controlled studies that deliver the same AV material with and without a production task, compare audio-visual with audio-only input, and use reliable ratings of fluency, pronunciation and accuracy with delayed follow-up.

## REFERENCES

- [1] Al-jumaily, I. H. A., & Alazzawi, I. T. J. (2025). The influence of multimodal visual methodologies on EFL university students' audio-visual comprehension, verbal and nonverbal communication. *Journal of Language & Education*, 11(2).†
- [2] Bahrani, T., & Sim, T. S. (2012). Audiovisual news, cartoons, and films as sources of authentic language input and language proficiency enhancement. *The Turkish Online Journal of Educational Technology*, 11(4), 56–64.
- [3] Bajrami, L., & Ismaili, M. (2016). The role of video materials in EFL classrooms. *Procedia – Social and Behavioral Sciences*, 232, 502–506. <https://doi.org/10.1016/j.sbspro.2016.10.068>
- [4] Campbell, M., McKenzie, J. E., Sowden, A., et al. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, 368, l6890.
- [5] Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–170.
- [6] Dewayanti, N. P. W., Artini, L. P., & Kusuma, I. P. I. (2026). Exploring the role of English movies in language learning: Senior high school students' perceptions. [Journal title to be confirmed], 5(8), 1322–1333.†
- [7] Fitri, R. D. (2022). *The effect of using animation video as audio-visual media toward students' speaking skill at the tenth grade of SMAN Plus Riau Province* [Undergraduate thesis, State Islamic University of Sultan Syarif Kasim Riau].
- [8] Hong, Q. N., Pluye, P., Fàbregues, S., et al. (2018). *Mixed Methods Appraisal Tool (MMAT), version 2018*. Canadian Intellectual Property Office.†
- [9] Irmawati, Ihsan, M., & Rasmi. (2020). Utilizing audio visual aids to improve English speaking skill for the eight grade students of SMP Negeri 3 Banawa. *Datokarama English Education Journal*, 1(1), 40–50.†
- [10] Kathirvel, K., & Hashim, H. (2020). The use of audio-visual materials as strategies to enhance speaking skills among ESL young learners. *Creative Education*, 11(12), 2599–2608. <https://doi.org/10.4236/ce.2020.1112192>
- [11] Kim, H.-S. (2015). Using authentic videos to improve EFL students' listening comprehension. *International Journal of Contents*, 11(4), 15–24.
- [12] Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- [13] Macwan, H. J. (2015). Using visual aids as authentic material in ESL classrooms. *Research Journal of English Language and Literature*, 3(1), 91–96.
- [14] Maulidiah, F., Kurrata'Aini, R., Yanti, M., & Martiarini, E. (2026). The effectiveness of using YouTube for teaching EFL in improving student's speaking skills: A systematic literature review. *Indonesian Research Journal on Education*, 6(2), 191–198.
- [15] Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press.
- [16] Mizab, M. (2015). *The role of audio-visual aids in enhancing EFL students' speaking skill* [Master's dissertation, Mohamed Kheider University of Biskra].
- [17] Naz, S. (2023). Role of multimedia-aided EFL classrooms in promoting learners' interaction and participation in tertiary-level Bangladeshi students. *Journal of Language Teaching and Research*, 14(5), 1207–1214. <https://doi.org/10.17507/jltr.1405.08>
- [18] Page, M. J., McKenzie, J. E., Bossuyt, P. M., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71.
- [19] Paivio, A. (1971). *Imagery and verbal processes*. Holt, Rinehart & Winston.
- [20] Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.
- [21] Pinto, M. A. M., & Lomi, A. N. K. (2026). Integrating task-based activities and audio-visual media to enhance EFL students' speaking ability. *Ebony*, 6(1), 78–97. <https://doi.org/10.37304/ebony.v6i1.23891>
- [22] Polat, M., & Erişti, B. (2019). The effects of authentic video materials on foreign language listening skill development and listening anxiety at different levels of English proficiency. *International Journal of Contemporary Educational Research*, 6(1), 135–154.
- [23] Rahman, M. M., & Jamila, M. (2024). Effectiveness of audiovisual materials in developing tertiary level learners' English listening and speaking skills. *International Journal of Language Testing*, 14(2), 67–81. <https://doi.org/10.22034/IJLT.2024.430050.1312>
- [24] Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., et al. (2021). PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Systematic Reviews*, 10, 39.
- [25] Saber, A. F. (2025). Examining the effectiveness of authentic video materials in developing speaking skills among undergraduate EFL Iraqi learners. *Journal of Tikrit University for Humanities*, 32(11–2), 22–37.
- [26] Saed, H. A., Haider, A. S., Al-Salman, S., & Hussein, R. F. (2021). The use of YouTube in developing the speaking

- skills of Jordanian EFL university students. *Heliyon*, 7(7), e07543.
- [27] Sabiq, A. H. (2017). Improving students' motivation and reading skills in ELT through audio visual media. *Diksi*, 25(1).† (Cited in text as "As Sabiq".)
- [28] Shaojie, T., Samad, A. A., & Ismail, L. (2022). Systematic literature review on audio-visual multimodal input in listening comprehension. *Frontiers in Psychology*, 13, 980133.
- [29] Shikongo, M. E. P. (2025). *Exploring teachers' use of audio-visual aids in teaching English-speaking skills to junior primary learners* [Master's thesis, University of Namibia].†
- [30] 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.
- [31] Syah, D. A., & Fediyanto, N. (2026). Animated videos in English listening skill development. *Academia Open*, 11(1).
- [32] Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.