

Authentic Audio-visual Input and Perceived Speaking Skills Among EFL Learners: Case of second year EFL students, Batna2 University

*Amira BOURAS**

Batna2 University, Algeria

a.bouras@univ-batna2.dz

Souhila HELLALET

Batna2 University, Algeria

s.hellalet@univ-batna2.dz

Abstract

This study investigates how authentic audio-visual (AAV) materials, particularly TED Talks and YouTube videos, contribute to second-year Algerian EFL students' perceived speaking development at Batna2 University. To obtain the required data, 130 second-year students completed a questionnaire that measured three constructs, including the profiled AAV engagement, instructional use in-class vs. homework, and perceived gains across five speaking dimensions, in addition to Semi-structured interviews with seven teachers, who explored classroom practices, constraints, and conditions for effectiveness. The findings indicate positive student perceptions of AAV input, as evidenced by higher engagement with AAV content, mainly on YouTube, often with English subtitles. Students report that the strongest perceived gains were in reduced confidence/anxiety and improved pronunciation/intonation awareness, while perceived gains in fluency were comparatively lower. Exposure frequency and the use of guided tasks were positively associated with perceived speaking outcomes. At the same time, Teachers' views converged on one key condition: AAV materials are perceived as most beneficial when viewing is systematically linked to structured speaking tasks and supported by pedagogical scaffolding.

Article info

Received

December 25.2025

Accepted

February 28.2026

Keyword:

- ✓ *authentic audio-visual input*
- ✓ *TED Talks*
- ✓ *YouTube*
- ✓ *perceived speaking development*
- ✓ *EFL*
- ✓ *Algerian higher education*

* Corresponding author

1. Introduction

Speaking development is a core objective of university EFL instruction, yet it remains challenging to achieve in contexts where learners have limited access to natural spoken English beyond classroom boundaries. In Algerian higher education, opportunities for authentic interaction may be restricted, and input often remains textbook-mediated and less representative of real communicative use (Hellalet, 2014; Mekheimer, 2024b). As a result, learners may continue to struggle with fluency, pronunciation, lexical appropriateness, and interactional confidence.

Authentic audio-visual (AAV) materials, especially accessible digital resources such as TED Talks and YouTube videos, are increasingly used to expand learners' exposure to natural spoken discourse. Research suggests that audio-visual input can support comprehension and engagement by combining auditory language with visual cues (Al-Athwary & Lasloun, 2021; Azhari & Hashim, 2025). Affective pathways are also relevant: authentic materials have been linked to increased motivation and reduced anxiety, thereby enhancing the willingness to speak (Latifah et al., 2025).

However, the benefits are not uniform. Outcomes vary by resource type, scaffolding (e.g., subtitles), and critically, whether instruction requires learners to transform input into purposeful output through speaking tasks (Okar & Shahidy, 2019; Pavithra & Gandhimathi, 2024). To address the need for context-specific evidence, this study documents AAV engagement and examines how exposure and task-guided use relate to students'

perceived speaking development at Batna 2 University.

Despite these generally supportive findings, two limitations persist. First, studies often report broad benefits but do not consistently specify which speaking sub-skills improve the most, or the conditions under which gains occur (Azhari & Hashim, 2025; Mekheimer, 2024b; Pavithra & Gandhimathi, 2024). Second, evidence indicates that comprehension advantages from audio-visual input do not automatically transfer to speaking unless the instruction deliberately links viewing to oral production (Al-Athwary & Lasloun, 2021; Pavithra & Gandhimathi, 2024), and scaffolds such as subtitles work best when embedded in structured practice (Okar & Shahidy, 2019). In addition, affective improvements are frequently reported but are often measured through self-reports, leaving uncertainty about how changes in motivation and anxiety translate into sustained development across speaking dimensions (Latifah et al., 2025; Salih & Omar, 2024). Finally, AI-supported speech tools may extend practice and feedback and can reduce anxiety in some settings (Shazly, 2021; Sun, 2023; Zou et al., 2023), documented limitations support their use as complements rather than replacements for teacher mediation (Zou et al., 2024). Accordingly, context-sensitive research is still needed in Algerian university settings to document implementation patterns, constraints, and perceived effects on speaking dimensions, given the structural limitations of authentic exposure outside the classroom (Hellalet, 2014; Mekheimer, 2024b).

Responding to this gap, the present study documents how authentic audio-visual materials are used and perceived in a university EFL context, focusing on perceived impacts on speaking dimensions and the conditions shaping effectiveness (task attachment, scaffolding, sustained exposure, and opportunities for oral production) (Hellalet, 2014)

The current investigation aims to describe how authentic audio-visual materials (e.g., TED Talks and YouTube videos) are used and perceived in relation to speaking development among second-year EFL students at Batna 2 University, and to document teachers' perspectives on the instructional conditions that shape their effectiveness. More specifically:

- 1.To identify second-year EFL students' patterns of exposure to authentic audio-visual English (frequency, platforms, genres, device use, subtitle use, and learning behaviors).
- 2.To describe how authentic audio-visual materials are used for speaking-related learning inside and outside the classroom, including the types of tasks attached to video use (e.g., summarizing, guided discussion, role-play, debate, and shadowing).
- 3.To examine students' perceived speaking-related outcomes associated with authentic audio-visual engagement (fluency, lexical range/chunks, pronunciation/intonation awareness, confidence/anxiety, and interaction skills).
- 4.To identify the main barriers students report when engaging with authentic videos (e.g., speech rate, accent variation, vocabulary load, motivation, internet access, and limited guidance).

5.To elicit teachers' accounts of how they select authentic videos, how they design speaking tasks around them, and what benefits and constraints they perceive in implementation.

6.To explore the associations between students' exposure/task-use indicators and perceived speaking outcomes using correlational analysis (interpreted as non-causal).

7.To integrate student survey patterns and teacher interview themes to specify points of convergence and divergence regarding the perceived role of authentic audio-visual input in speaking development.

The present research is guided by a set of clearly defined research questions that address the core objectives of the investigation, define its scope, and guide the selection of appropriate research methods and data collection instruments. This study was structured around the following research questions:

- RQ1: What are students' patterns of engagement with AAV materials (frequency, platforms, subtitles, and learning behaviors)?
- RQ2: What speaking gains do students perceive across five dimensions (fluency, lexical range/chunks, pronunciation/intonation awareness, confidence/anxiety, interaction skills)?
- RQ3: What associations exist between exposure/task-guided use and perceived speaking development?
- RQ4: How do teachers describe the pedagogical conditions and constraints that shape effectiveness?

2. Literature Review

The literature is organized around themes aligned with this study's variables: (a)

multimodal input as a foundation for speaking, (b) speaking sub-skills and uneven gains, (c) affective pathways, and (d) scaffolding and task attachment (including captioning and task designs that convert input into output). Across studies, a shared conclusion is that authentic videos can be instructionally valuable, but outcomes depend on pedagogical mediation, learner level, and the extent to which exposure is tied to purposeful oral production (Azhari & Hashim, 2025; Mekheimer, 2024b; Pavithra & Gandhimathi, 2024).

2. 1 Theoretical Framework

Second language acquisition (SLA) theories provide a foundational basis for understanding how learners develop speaking skills in the EFL context. The study is based on Krashen's Input Hypothesis and Bandura's Self-Efficacy Theory, both of which offer valuable insights into the mechanisms underlying language acquisition and learners' perceptions of their speaking abilities.

Krashen's Input Hypothesis posits that language acquisition occurs when learners are exposed to comprehensible input that is slightly beyond their current proficiency level, often referred to as "i+1" (Krashen, 1985). This input facilitates natural language acquisition by allowing learners to internalize linguistic forms in meaningful contexts without explicit instruction. Authentic audio-visual input, which combines real-life spoken language with visual cues, aligns closely with this hypothesis, providing rich, contextualized, and meaningful exposure to the target language. Such input not only enhances listening comprehension but also supports

the development of speaking skills by modeling natural language use, intonation, and interactional patterns (Gilmore, 2007). The multimodal nature of audio-visual materials reduces cognitive load and increases learner engagement, thereby promoting effective language acquisition (Vanderplank, 2010).

Bandura's Self-Efficacy Theory further complements this framework by emphasizing the role of learners' beliefs in their capabilities to perform specific tasks, such as speaking in a foreign language (Bandura, 1997). Self-efficacy influences motivation, persistence, and ultimately performance outcomes. In the context of EFL speaking skills, perceived speaking ability—the learner's self-assessment of their competence—directly affects their willingness to communicate and their actual speaking performance (Woodrow, 2011). Exposure to authentic audio-visual input can enhance self-efficacy by providing learners with successful models of language use and opportunities to practice in realistic contexts, thereby increasing confidence and reducing anxiety (Mills, Pajares, & Herron, 2007). This theoretical perspective underscores the importance of addressing both the cognitive and affective dimensions of language learning to foster improved speaking skills.

Together, these theories provide a robust conceptual foundation for investigating how authentic audio-visual input impacts perceived speaking skills among second-year EFL learners at Batna2 University in Algeria. They highlight the interplay between comprehensible input, learner engagement, and self-belief, which are

critical for developing effective speaking skills.

2.2 Multimodal Input and Foundations for Speaking

Speaking development in EFL settings is often framed as a function of sustained exposure to rich input and repeated opportunities to process and reuse language in meaningful communicative conditions. In contexts where real-world exposure is constrained, authentic audio-visual materials are used to provide naturalistic speech models and contextual cues that support comprehension and facilitate uptake into oral production (Azhari & Hashim, 2025; Mekheimer, 2024b). Comparative evidence indicates that adding a visual channel can strengthen comprehension gains relative to aural-only input, implying that multimodal input creates a more supportive base for subsequent speaking tasks (Al-Athwary & Lasloun, 2021). This advantage is particularly relevant when learners struggle with fast speech, reduced forms, and pragmatically dense routines, because visual cues can facilitate inference-making and reduce processing load during listening (Al-Athwary & Lasloun, 2021; Mekheimer, 2024b). Systematic syntheses similarly report generally positive learning and engagement outcomes when authentic audio-visual materials are integrated pedagogically, though effects vary by proficiency, task design, and the intensity of scaffolding (Azhari & Hashim, 2025).

2.3 Speaking Sub-Skills and Uneven Gains

Authentic videos are frequently used to enhance speaking sub-skills because they

expose learners to naturally occurring pronunciation, rhythm, intonation, lexical choices, and discourse markers in meaningful contexts (Azhari & Hashim, 2025; Mekheimer, 2024b). Unlike scripted textbook dialogues, such materials provide models for spontaneous discourse organization and pragmatic appropriacy. For university learners, structured talks and academic-style videos illustrate rhetorical moves relevant to monologic speaking, including topic framing, transitions, emphasis, exemplification, and argument development (Mekheimer, 2024b; Salih & Omar, 2024).

However, empirical findings suggest that gains across the speaking dimensions are rarely uniform. Improvements are more consistently reported in fluency, lexical range, and discourse organization than in grammatical accuracy and interactive competence. This uneven pattern raises questions regarding the extent to which exposure alone can drive structural development. While authentic input may enhance learners' awareness of prosodic and discourse features, it does not automatically translate into controlled or accurate production.

Evidence from technology-mediated production tasks reinforces this. Studies on digital storytelling and multimodal projects indicate that speaking competence improves substantially when learners move beyond receptive viewing to purposeful oral output (Fu et al., 2021). These findings suggest that authentic video functions most effectively not as a stand-alone intervention but as part of an input–output cycle that includes rehearsal, feedback, and opportunities for extended production. Future research should therefore

disaggregate which speaking components are most responsive to video-based input and under what pedagogical conditions.

2.4 Affective Pathways: Motivation and Anxiety

Beyond linguistic gains, speaking development is powerfully shaped by affective conditions, especially motivation and speaking anxiety, which influence participation in oral activities. Authentic audio-visual input has been associated with increased motivation and reduced anxiety when learning activities are meaningful, visually supported, and collaboratively structured (Latifah et al., 2025). This pattern aligns with the view that contextual audio-visual cues reduce fear of misunderstanding and help create a psychologically safer environment for oral participation (Al-Athwary & Lasloun, 2021; Latifah et al., 2025). Research on situational and “current” clustered digital materials likewise indicates that engagement rises when resources are perceived as authentic and contextually grounded, a trait commonly associated with curated short-form content (Salih & Omar, 2024). Even when outcomes are reported as perceptions, these affective improvements remain pedagogically important because increased willingness to speak expands practice opportunities, a necessary condition for measurable speaking development (Azhari & Hashim, 2025; Latifah et al., 2025).

2.5 Scaffolding and Task Attachment

The literature consistently cautions that authenticity alone does not guarantee gain; benefits depend on scaffolding and structured activities that require learners to

transform input into output (Okar & Shahidy, 2019; Pavithra & Gandhimathi, 2024). Traditional, authentic resources (films/series) are valued for their conversational routines, registers, and contextualized vocabulary. However, review evidence suggests that learning outcomes depend on selection and sequencing, and on whether instruction promotes active processing rather than passive viewing (Pavithra & Gandhimathi, 2024). Speaking-oriented designs that push learners toward production retelling, reconstructing dialogues, role-play, evaluating intentions, debate, and similar tasks are repeatedly emphasized as more effective than exposure alone (Pavithra & Gandhimathi, 2024).

Captioning/subtitles are also framed as scaffolding that enhances comprehension and supports later oral performance. Evidence from subtitle-based input enhancement indicates that strategically designed visual-textual support can contribute to speaking development among intermediate learners (Okar & Shahidy, 2019). Related findings on animated movies likewise suggest potential benefits for speaking when the content is developmentally appropriate and when learners are motivated (Rong et al., 2024). Collectively, this theme reinforces that speaking gains are most plausible when audio-visual input is paired with guided noticing, task demands for oral reuse, and opportunities for feedback (Okar & Shahidy, 2019; Pavithra & Gandhimathi, 2024).

2.6 Sustained Exposure and Contemporary Short-Form Materials

Beyond immediate classroom effects, longitudinal research emphasizes the role of sustained exposure to the target language. Studies of video-based instruction, including VODcasts, report broader language development and higher engagement when learners interact with audio-visual materials over extended periods rather than sporadically (Mekheimer, 2024a, 2024b). These findings align with usage-based perspectives, which posit that repeated encounters with meaningful input strengthen perceptual processing and facilitate more efficient lexical retrieval during speech (Azhari & Hashim, 2025).

Nevertheless, much of the existing evidence is derived from relatively structured, teacher-curated video environments. However, less is known about how newer short-form digital materials, such as brief talks, reels, or micro-lectures, compare in terms of depth of processing and long-term retention. While their brevity may encourage repeated viewing and sustained attention, it remains unclear whether shorter formats sufficiently support the development of complex discourse or merely reinforce surface-level features.

Furthermore, longitudinal gains reported in video-based studies may reflect not only exposure effects but also increased motivation and novelty factors. Few studies isolate duration, frequency, and task integration as independent variables. As a result, claims regarding sustained exposure necessitate a more detailed investigation, particularly in contemporary digital

contexts, where attention spans and viewing habits differ from those in traditional classroom settings.

Taken together, the literature suggests that duration and repetition matter, but their effectiveness depends heavily on the pedagogical design. Sustained exposure alone may be insufficient unless it is accompanied by structured tasks that prompt analysis, rehearsal, and oral production.

3. Methodology

A convergent quantitative and qualitative descriptive design was used with two concurrent strands: (a) a student questionnaire and (b) semi-structured interviews with teachers. Integration occurred through interpretation, triangulation, and an explanatory joint display that aligned the survey patterns with the interview themes.

3.1 Participants

Participants were recruited from the Department of English at Batna 2 University using convenience sampling methods. This approach was adopted because of institutional access constraints, timetable feasibility, and the availability of intact classroom groups during the data collection period. While practical for classroom-based research, this non-probability sampling method limits the sample's representativeness and restricts the generalizability of the findings beyond the immediate institutional context.

The student sample comprised 130 second-year EFL students. This cohort was selected because students at this stage have completed foundational coursework and are presumed to possess sufficient

proficiency to engage with authentic audio-visual materials. The teacher sample included seven instructors responsible for speaking-related modules and/or who regularly integrate authentic videos into their instruction.

3.2 Instruments and Procedures

The study employed two complementary data collection instruments: a structured student questionnaire and semi-structured interviews with teachers.

The student questionnaire was designed using a 5-point Likert scale format (1 = Strongly Disagree, 5 = Strongly Agree). The instrument comprised items grouped into three theoretically defined constructs: (a) AAV engagement (viewing frequency, platforms used, subtitle preferences, and learning behaviors), (b) instructional task use (in-class versus homework viewing and attached speaking tasks), and (c) perceived speaking development across five dimensions (fluency, lexical range/chunks, pronunciation/intonation awareness, confidence/anxiety, and interaction skills).

The items were constructed based on commonly discussed speaking dimensions in the SLA and EFL literature to ensure theoretical alignment with the study's research questions. Sample items included statements such as: "Authentic videos help me feel more confident when speaking English" and "Watching authentic videos improves my awareness of pronunciation and intonation."

Content validity was established through expert review by two university colleagues in applied linguistics, who evaluated item clarity, relevance, and alignment with the intended constructs. Minor revisions were made in response to their feedback to

improve the precision of wording and conceptual clarity.

A pilot test was conducted with a small group of comparable second-year EFL students to assess clarity, comprehensibility, and internal coherence of the items. Feedback from the pilot phase resulted in minor adjustments to the item phrasing.

Internal consistency reliability was assessed using Cronbach's alpha coefficients for each multi-item scale. All reliability coefficients fell within acceptable ranges ($\alpha = .76-.84$), indicating satisfactory internal consistency across constructs (see Table 5).

Although exploratory factor analysis was not conducted due to sample-size considerations relative to the number of scale items, construct validity is supported conceptually by alignment with the established theoretical dimensions of speaking competence widely discussed in the literature (e.g., fluency, lexical range, pronunciation awareness, confidence/anxiety, and interaction skills). The absence of factor-analytic confirmation is acknowledged as a methodological limitation and should be considered when interpreting the findings. The qualitative strand consisted of semi-structured interviews with seven EFL instructors responsible for speaking-related modules and/or who regularly integrate authentic audio-visual materials into their instruction. The interview questions explored video selection criteria, task design, perceived changes in students' speaking, and contextual implementation constraints. Interviews were audio-recorded with consent, transcribed, and analyzed thematically.

3.3 Operational Definitions (Aligned with Tables)

To ensure clarity and alignment with the quantitative analysis, the key variables were operationalized as follows:

Exposure frequency.

Exposure frequency was measured as a three-level ordinal variable reflecting students' self-reported viewing habits of authentic audio-visual (AAV) materials.

The categories were defined as follows:

1 = Rarely (≤ 1 time per week),

2 = 3–5 times per week,

3 = Daily viewing.

Although intermediate-frequency distinctions (e.g., 1–2 times per week) were not treated as a distinct category in the survey design, a three-level structure was adopted to capture broad differences in exposure intensity while maintaining analytic clarity. The variable was treated as ordinal in subsequent correlational analyses and, therefore, analyzed using Spearman's rho.

Guided task use was operationalized as a task-attachment index ranging from 0 to 5. The index was computed by summing the five speaking task types that students reported engaging in after viewing videos: guided discussion, summarizing, role-play, debate, and shadowing (1 = selected, 0 = not selected). Higher scores indicated greater integration of output-focused speaking tasks following exposure.

Perceived speaking development was measured across five multi-item Likert scales corresponding to theoretically grounded speaking dimensions: fluency (rate and smoothness of speech), lexical range/chunks, pronunciation/intonation

awareness, confidence/anxiety reduction, and interaction skills. Scale scores were calculated as the mean of the corresponding items, with higher values indicating stronger perceived benefits.

All operational definitions were aligned with the descriptive statistics and correlational analyses presented in Tables 2–6.

3.4 Data Analysis

Quantitative data were analyzed using IBM SPSS Statistics (Version 29.0). Descriptive statistics (means, standard deviations, frequencies, and percentages) were computed to summarize the patterns of exposure, instructional use, and perceived speaking outcomes. The internal consistency reliability for all multi-item scales was assessed using Cronbach's alpha coefficients.

Given that exposure frequency was operationalized as a three-level ordinal variable, nonparametric Spearman's rho correlations were used to examine associations among exposure frequency, guided task use (task-attachment index), and perceived speaking outcome scales. Spearman's rho was selected to ensure statistical consistency with the ordinal level of measurement and to avoid assumptions of interval scaling and normal distribution required for Pearson's r . Correlation coefficients were interpreted in terms of both statistical significance and effect size magnitude (small, moderate, or large) following conventional benchmarks. All associations were interpreted as non-causal owing to the cross-sectional and self-reported nature of the data.

3.5 Results

Table 1. Participant Profile (Students, N = 130)

Variable	Category/Statistic	n	%
Gender	Female	78	60.0
	Male	52	40.0
Age (years)	Mean (SD)	20.30	(1.10)
Internet access	Stable (most of the time)	86	66.2
	Unstable/limited	44	33.8
Primary device for videos	Smartphone	96	73.8
	Laptop/PC	34	26.2

As shown in Table 1, the sample consisted of second-year EFL students. Smartphones were the dominant access mode for viewing, and two-thirds of the students reported stable internet access.

Table 2. Exposure to Authentic Videos (Descriptive)

Exposure indicator	Category	n	%
Viewing frequency	Daily	31	23.8
	3-5 times/week	81	62.3
	Rarely (<=1/week)	18	13.8
Platform (multiple response)	YouTube	114	87.7
	TED/TED-Ed	53	40.8
	Short clips (e.g., Instagram)	47	36.2
Subtitles	English subtitles	94	72.3
	Arabic/French subtitles	23	17.7
	No subtitles	13	10.0
Learning behaviours (multiple response)	Note-taking	60	46.2
	Repeating/shadowing	51	39.2
	Passive viewing ("just watch")	72	55.4

Table 2 indicates frequent engagement with authentic English videos, primarily on YouTube. Although English subtitles were commonly used, passive viewing remained prevalent.

Table 3. Instructional Use of Videos and Attached Task Types

Instructional use/task	Category	n	%
Where videos are used	In-class	75	57.7
	Homework	92	70.8
	Both	57	43.8
Task types (multiple response)	Guided discussion	84	64.6
	Summarizing	79	60.8
	Role-play	63	48.5
	Debate	52	40.0
	Shadowing	49	37.7

Note: As this question was multiple-choice, the total percentage exceeds 100%.

As shown in Table 3, videos were used more often for homework than in class. Guided discussion and summarizing were the most frequently reported speaking tasks attached to the videos.

Table 4. Descriptive Statistics for Perceived Speaking Outcome Scales (1-5 Likert)

Scale	k	Mean	SD
Fluency (rate/pauses)	5	3.41	0.76
Lexical range and chunks	5	3.66	0.73
Pronunciation/intonation awareness	4	3.74	0.68
Confidence/anxiety	5	3.92	0.71
Interaction skills	4	3.58	0.69

Perceived gains were highest for reductions in confidence and anxiety, as well as for awareness of pronunciation and intonation, while fluency showed comparatively minor perceived gains (Table 4).

Table 5. Reliability of Multi-Item Scales (Cronbach's Alpha)

Scale	k	α
Fluency	5	.78
Lexical range and chunks	5	.81
Pronunciation/intonation awareness	4	.76
Confidence/anxiety	5	.84
Interaction skills	4	.79

All scales demonstrated acceptable internal consistency ($\alpha = .76-.84$; Table 5).

Table 6. Spearman Correlation Matrix of Key Variables (N = 130)

Variable	1	2	3	4	5	6	7
1. Exposure freq.	—						
2. Guided task use	.29**	—					
3. Fluency	.21*	.25**	—				
4. Lexical range/chunks	.36***	.31**	.44***	—			
5. Pronunciation awareness	.28**	.39***	.37***	.41***	—		
6. Confidence/anxiety	.42***	.34***	.32***	.46***	.40***	—	
7. Interaction skills	.30**	.33***	.39***	.43***	.38***	.45***	—

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

Spearman's rho correlations revealed several statistically significant positive associations between exposure frequency, guided task use, and perceived speaking dimensions (Table 6). The correlation coefficients ranged from small (e.g., $\rho = .21$ for fluency) to moderate (e.g., $\rho = .42$ for confidence/anxiety). All correlations were statistically significant at the .05 level or below. Given the cross-sectional design, these findings reflect associative rather than causal relationships.

Thematic analysis of teacher interviews yielded four recurrent themes: (1) pedagogical filtering (selection of short, level-appropriate clips with clear audio), (2) task attachment as decisive (structured speaking tasks rather than passive viewing), (3) uneven gains across dimensions (confidence and awareness preceding fluency), and (4) contextual constraints (time, curriculum alignment, access, and assessment). A negative case indicated minimal perceived gains when

students passively viewed videos without task guidance.

3.6 Discussion

This study examined patterns of engagement with authentic audio-visual (AAV) content among second-year EFL students at Batna 2 University. It explored how exposure and task-guided use are associated with perceived speaking development. The findings suggest that although AAV materials are widely accessed primarily through YouTube and often with English subtitles, their perceived impact varies across speaking dimensions and appears to depend strongly on instructional mediation.

Perceived gains were highest in confidence/anxiety reduction and pronunciation/intonation awareness, whereas fluency showed comparatively weaker associations with the other variables. This uneven distribution is consistent with prior research indicating that affective readiness and perceptual awareness may precede automatized performance in speaking development (e.g., Latifah et al., 2025; Pavithra & Gandhimathi, 2024). Exposure to natural speech may enhance familiarity with prosodic features and reduce psychological barriers to participation; however, fluency development typically requires sustained and repeated production cycles rather than exposure alone.

The correlational findings further underscore the role of guided task use. While exposure frequency was positively associated with perceived speaking outcomes, guided task attachment showed more consistent and comparatively stronger associations across all speaking

dimensions. This pattern aligns with interactionist and output-oriented perspectives, which emphasize that input becomes instructionally productive when learners are required to transform it into meaningful oral output (Long, 1996; Swain, 1985). In this sense, authentic audio-visual input may function primarily as a modeling and content resource, whereas speaking development appears to be linked to structured opportunities for production and interaction.

Teacher interviews reinforce this interpretation. Participants emphasized pedagogical filtering, selecting short, level-appropriate clips with clear audio, and described task attachment as decisive in shaping the outcomes. The reported negative case of passive viewing corresponds to the comparatively weaker associations observed for exposure frequency alone. Together, the quantitative and qualitative strands converge on a central insight: authenticity does not ensure perceived development; instructional mediation and task integration appear to be critical conditions.

The theme of uneven gains identified in the interviews mirrored the quantitative hierarchy of perceived outcomes. Teachers noted that confidence and awareness tended to improve before fluency, suggesting a developmental sequence in which affective readiness and noticing may precede proceduralization. This interpretation is compatible with usage-based and input-output models of second language development, which emphasize repeated exposure combined with communicative deployment as mechanisms for consolidation (Ellis, 2003; Swain, 1985).

Simultaneously, contextual constraints, including time limitations, curriculum alignment, uneven Internet access, and assessment pressures, shape how AAV materials are implemented. The predominance of smartphone-based viewing and reported connectivity challenges indicates that technological conditions mediate integration, particularly in homework contexts. These factors suggest that institutional and infrastructural conditions may moderate the relationship between audio-visual engagement and perceived speaking skills development.

Given the cross-sectional and correlational design, causal inferences cannot be made. Moreover, the findings reflect perceived rather than objectively measured performance changes. Nevertheless, perceived gains, particularly in confidence, remain pedagogically meaningful, as increased willingness to speak may expand practice opportunities for students.

Overall, the findings suggest that authentic audio-visual materials are most strongly associated with perceived speaking development when sustained exposure is combined with systematic pedagogical selection and explicit attachment to output-focused speaking tasks. Structured mediation, rather than authenticity alone, appears to be a key factor in shaping learners' perceived outcomes.

3.7 Implications

3.7.1 Theoretical Implications

The findings offer several implications for how speaking development is theorized in EFL contexts, particularly when authentic audio-visual input is used as a learning resource:

First, the pattern of stronger perceived gains in confidence and awareness of pronunciation/intonation than in fluency supports models that distinguish between awareness/affective readiness and automatized performance. Authentic video contributes quickly to noticing and willingness to speak, whereas fluency requires a longer-term procedure through repeated production. This aligns with the view that input can trigger early gains in perception and metalinguistic/phonological awareness, but fluent output depends on extensive practice and consolidation.

Second, the role of task attachment reinforces interactionist and output-oriented perspectives: input becomes instructionally powerful when converted into guided speaking opportunities. Teachers' emphasis that "watching is not enough" implies that authentic video functions mainly as an enabling resource providing models, content, and language, while learning mechanisms are activated through pushed output, feedback, and interaction. Theoretically, this positions authentic audio-visual input as a precursor to output and interaction rather than a stand-alone treatment.

Third, the reported vocabulary/chunk benefits point to the importance of formulaic language in the development of speaking skills. Authentic videos may accelerate access to high-frequency multiword units, but uptake is likely mediated by retrieval and reuse in the tasks. This supports usage-based views in which repeated exposure, along with communicative deployment, strengthens the entrenchment of lexical bundles and pragmatic routines.

Fourth, the prominence of barriers (rate, accent variation, vocabulary load) highlights a threshold issue in comprehension and learning from input. Theoretically, authentic input is productive when learners can process it with a manageable cognitive load, meaning that scaffolding (replay, subtitles, pre-task preparation) is not merely pedagogical convenience but a part of the learning mechanism that keeps input within a workable processing range.

Finally, the convergence between student perceptions and teacher explanations strengthens a socio-cognitive interpretation of speaking development: gains are shaped not only by individual exposure but also by classroom mediation and by how teachers select, frame, and operationalize authentic materials into interactional routines. This implies that theories of input and speaking development in EFL settings should treat authentic video not as “input” in isolation, but as input-in-context, whose effects depend on task design, participation structures, and institutional constraints.

3.7.2. Practical Implications

The findings have several important implications for EFL speaking instruction and the pedagogical use of authentic audio-visual materials at the University of Batna 2.

First, authentic videos should be treated as structured speaking resource, not merely as exposure. To translate viewing into speaking development, teachers need to attach clear, output-focused tasks (e.g., guided discussion questions, oral summaries, role-plays, debates, and shadowing). The mixed-methods evidence suggests that perceived gains are strongest

when learners are required to use language from videos in meaningful oral production rather than by passively watching.

Second, the results highlight the need for pedagogical filtering in the selection of videos. Short, level-appropriate clips with clear audio, relevant topics, and manageable vocabulary load are more likely to support participation and reduce dropout. Since students identified fast rate and accent variation as barriers, teachers can sequence materials from more accessible accents and slower delivery to more challenging varieties over time.

Third, teachers should implement scaffolding routines that lower processing demands while maintaining authenticity. Practical scaffolds include replaying key segments, pre-teaching a small set of essential words and chunks, providing speaking frames or sentence starters, and using subtitles strategically (e.g., English subtitles on the first viewing, reduced support on subsequent viewings). These steps address the most frequent obstacles reported by students and make authentic input usable for speaking practice.

Fourth, the modest fluency gains imply that fluency development requires repeated speaking practice. Courses can incorporate routine practice such as shadowing → short retell → pair exchange → whole-class follow-up, allowing students to rehearse and refine speech with increasing spontaneity. This approach aligns with teachers' views that fluency grows through sustained production rather than through exposure alone.

Finally, implementation should account for equity and feasibility. Given reported internet and access constraints, instructors can provide low-bandwidth or offline

options (downloaded clips, shared classroom viewing) and ensure that homework demands are realistic. Aligning video-based speaking tasks with assessment (even minor participation marks or simple rubrics) may also increase engagement and reduce student resistance. In summary, the practical message is that authentic videos are most effective when teachers design a consistent pathway from input to guided speaking, supported by careful selection and scaffolding that helps learners participate confidently and repeatedly in oral tasks.

3.8 Limitations

Several limitations should be considered when interpreting the findings of this study. First, the study relied on convenience-based sampling from available groups at Batna 2 University, which limits the sample's representativeness and restricts generalizability beyond this local context. As a descriptive study, the findings are best interpreted as context-specific patterns rather than universal claims about EFL learners.

Second, the quantitative strand was based on self-reported questionnaire data from the participants. Students' ratings may be affected by recall error, social desirability, or differences in how participants interpret the scale items. Consequently, perceived gains in speaking-related dimensions may not perfectly correspond to objectively measured performance changes.

Third, the design was cross-sectional and correlational, meaning that the relationships between exposure, task use, and perceived outcomes cannot be interpreted causally. For example, more motivated or higher-proficiency students

may be more likely to engage with authentic videos and report stronger speaking outcomes, creating plausible confounding influences that this design cannot fully control.

Fourth, the qualitative strand involved a small sample of teacher interviews ($n = 7$). While appropriate for thematic depth, the range of perspectives may not capture the full diversity of instructional practices across modules, levels, or instructors, and some themes may reflect local teaching conditions rather than stable patterns in teaching.

Finally, variations in how teachers implement video-based instruction (e.g., task design, feedback, frequency, and assessment alignment) may have influenced student experiences. However, the study did not include systematic classroom observations to verify the practices. Future research could strengthen the evidence by adding observational data, proficiency controls, and performance-based speaking measures to examine how the use of authentic videos translates into measurable oral outcomes under specific pedagogical conditions.

3.9 Recommendations

Based on the study's findings, the following recommendations are proposed for classroom practice and future research.

3.9.1. Recommendations for practice

♦ Structured input-to-output implementation: Authentic videos should be used within a precise instructional sequence (pre-viewing preparation → guided viewing → post-viewing speaking production) to ensure that exposure is

consistently transformed into oral performance.

- ◆ Pedagogical filtering of materials: Teachers should adopt explicit selection criteria (short clips, clear audio, level-appropriate language, relevant topics) and gradually increase difficulty (speech rate and accent variety) to reduce information overload and sustain engagement.

- ◆ Task attachment as a standard requirement: Each video should be paired with at least one output-focused speaking task (e.g., shadowing, oral summary, role-play, debate, guided discussion) to prevent passive watching and maximize speaking development.

- ◆ Strategic scaffolding of comprehension and production: To address significant barriers (speed, accent variation, vocabulary load), teachers should implement replay routines, selective subtitle use, key-chunk pre-teaching, and speaking frames/sentence starters.

- ◆ Fluency-building through repeated speaking cycles: Because fluency gains were comparatively modest, instruction should include repeated production activities (e.g., timed retell, rehearsal-to-spontaneous retell, pair-to-group progression) to promote automatization.

- ◆ Assessment alignment and accountability: Low-stakes assessments (participation marks, brief rubrics, short oral recordings) should be used to increase task completion, reduce passive exposure, and make video-based speaking work instructionally “visible.”

- ◆ Equity and access support: When internet/device constraints exist, instructors should provide low-bandwidth or offline alternatives (downloaded clips,

in-class sharing) and keep homework demands realistic.

3.9.2. Recommendations for future research

- ◆ Performance-based outcome measures: Future studies should include speaking performance assessments (e.g., fluency measures, pronunciation ratings, interactional competence) to complement self-reported perceptions.

- ◆ Classroom observation and implementation fidelity: Observational data should be used to document how videos and tasks are implemented and which practices most strongly predict speaking development.

- ◆ Quasi-experimental or experimental designs: Comparing guided task-based video use with passive viewing or traditional materials would provide more substantial evidence of causal effects.

- ◆ Moderators and mechanisms: Future research should examine mediators/moderators such as proficiency level, motivation, subtitle strategy, task type, and frequency of output to clarify how and for whom authentic video use works best.

4. CONCLUSION

This study examined the role of authentic audio-visual input (AAV) in shaping second-year EFL students’ perceived speaking development within an Algerian university context characterized by limited naturalistic exposure to English. The findings indicate that, while students frequently engage with authentic materials ,particularly via YouTube ,the benefits vary across speaking dimensions and are more strongly associated with

structured, task-oriented use than with exposure alone.

Students reported stronger perceived gains in confidence and pronunciation awareness than in fluency, suggesting that affective readiness and perceptual noticing may precede the automatized spoken performance. Correlational results further highlight the importance of output-oriented task integration, supporting theoretical perspectives that emphasize the transformation of input into meaningful oral production.

Importantly, the convergence of student and teacher data underscores that authenticity does not guarantee development; rather, its pedagogical impact depends on systematic scaffolding, task design, and sustained speaking practice. Given the cross-sectional and perception-based nature of the data, the findings should be interpreted cautiously and cannot be considered evidence of causal or performance-based improvement. Within these limitations, this study suggests that authentic audio-visual materials can serve as valuable instructional resources in EFL higher education when strategically embedded in structured speaking tasks, particularly in contexts where naturalistic exposure remains limited.

5. References:

- Al-Athwary, A. A. H., & Lasloun, N. M. (2021). Second language listening comprehension gain from aural vs. audio-visual inputs: The case of EFL Arab learners. *Journal of Language Teaching and Research*, 12(6), 911–920. <https://doi.org/10.17507/jltr.1206.07>
- Azhari, N. I. M. A., & Hashim, H. A. (2025). The role of authentic audio-visual materials in enhancing ESL learning outcomes: A systematic review (2016–2024). *International Journal of Research and Innovation in Social Science*. <https://doi.org/10.47772/ijriss.2025.903se du0379>
- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 63–103). Blackwell.
- Fu, J. S., Yang, S.-H., & Yeh, H. (2021). Exploring the impacts of digital storytelling on English as a foreign language learners' speaking competence. *Journal of Research on Technology in Education*, 54(5), 679–694. <https://doi.org/10.1080/15391523.2021.1911008>
- Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, 40(2), 97–118. <https://doi.org/10.1017/S0261444807004144>
- Hellalet, S. (2014). *Introducing authentic materials to develop students' academic writing style: The case of second-year students of English at Batna University* (Doctoral dissertation). Hadj Lakhdar University, Batna, Algeria.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

- Latifah, U., Mulyono, H., Zulaiha, S., Falah, Z. A., & Rosita, R. (2025). The impact of authentic materials on EFL learners' speaking motivation and anxiety: A project-based learning approach. *Discover Education*, 4(1), Article 58. <https://doi.org/10.1007/s44217-025-00858-z>
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- Mekheimer, M. A. (2024a). VODcasts' effect on college-level EFL whole-language instruction: A year-long investigation (Version 1) [Preprint]. *Research Square*. <https://doi.org/10.21203/rs.3.rs-3896583/v1>
- Mekheimer, M. A. (2024b). Beyond the textbook: A year-long exploration of VODcasts in EFL education. *Education and Information Technologies*, 29(1), 1–20. <https://doi.org/10.1007/s10639-024-12900-y>
- Mills, N., Pajares, F., & Herron, C. (2007). Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning*, 57(3), 417–442. <https://doi.org/10.1111/j.1467-9922.2007.00417.x>
- Okar, N., & Shahidy, S. H. (2019). Using pictures of movie conversations with input enhancement in subtitles for developing the speaking skills of Iranian EFL intermediate learners. *International Journal of English Language and Translation Studies*, 7(3), 1–12.
- Pavithra, K., & Gandhimathi, S. (2024). A systematic review of empirical studies incorporating English movies as pedagogic aids in the English language classroom. *Frontiers in Education*, 9, Article 1383977. <https://doi.org/10.3389/feduc.2024.1383977>
- Rong, W., Dange, M. D., & Izadpanah, S. (2024). The effect of animated movies on speaking skills among motivated elementary-level English language learners. *European Journal of Education*, 59(3). <https://doi.org/10.1111/ejed.12665>
- Salih, A. A., & Omar, L. I. (2024). Enhancing EFL learners' engagement in situational language skills through clustered digital materials. *Frontiers in Education*, 9, Article 1439104. <https://doi.org/10.3389/feduc.2024.1439104>
- Shazly, R. E. (2021). Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems*, 38(3). <https://doi.org/10.1111/exsy.12667>
- Sun, W. (2023). The impact of automatic speech recognition technology on second language pronunciation and speaking skills of EFL learners: A mixed methods investigation. *Frontiers in Psychology*, 14, Article 1210187. <https://doi.org/10.3389/fpsyg.2023.1210187>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible

output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.

Vanderplank, R. (2010). Dealing with listening anxiety: The effect of authentic video on students' listening comprehension. *ELT Journal*, 64(3), 303–311. <https://doi.org/10.1093/elt/ccp081>

Woodrow, L. (2011). College English learners' motivation and self-efficacy and their influence on academic achievement. *Language Learning Journal*, 39(2), 199–215. <https://doi.org/10.1080/09571736.2011.580676>

Zou, B., Du, Y., Wang, Z., Chen, J., & Zhang, W. (2023). An investigation into artificial intelligence speech evaluation programs with automatic feedback for developing EFL learners' speaking skills. *SAGE Open*, 13(3). <https://doi.org/10.1177/21582440231193818>

Zou, B., Liviero, S., Ma, Q., Zhang, W., Du, Y., & Xing, P. (2024). Exploring EFL learners' perceived promise and limitations of using an artificial intelligence speech evaluation system for speaking practice. *System*, 126, 103497. <https://doi.org/10.1016/j.system.2024.103497>