

TEACHER GUIDED YOUTUBE INTEGRATION IN BANGLADESHI FOUNDATION ENGLISH CLASSES

MD. ALI REZWAN TALUKDAR¹, MD. AMAZ UDDIN CHOWDURY², FATIMA YEASMIN CHANDHA²

¹ Faculty of English, Comilla University, Cumilla, Bangladesh

² Faculty of English, Hamdard University Bangladesh, Gazaria, Munshiganj-1510

rezwan@cou.ac.bd; amazuddin@hamdarduniversity.edu.bd; fyc@hamdarduniversity.edu.bd

First Received: 10 August. 2026

Final Proof Received: 20 Oct. 2026

ABSTRACT

This study examines teacher guided YouTube integration in Foundation English speaking classes in Bangladesh, focusing on instructional practices, perceived benefits, and constraints on participation. An exploratory descriptive design combined questionnaires from 103 first-year undergraduates at a private university, interviews with five English language teaching faculty members, and four classroom observations. Sixteen videos were introduced across twelve speaking classes. The analysis integrates questionnaire summaries, faculty perspectives, and observation episodes. Three survey figures present reported percentages, with unresolved discrepancies limiting quantitative interpretation. Teachers connected video input with speaking through vocabulary preparation, explanation during viewing, and subsequent pair or group activities. Participants described confidence, pronunciation models, and subtitles as useful; observations documented voluntary contributions, follow-up questions, and reuse of expressions. However, demanding language, unfamiliar cultural references, unequal speaking opportunities, connectivity interruptions, and limited responsive feedback constrained participation. These findings identify teacher mediation as a condition of the classroom experience. They do not demonstrate gains in speaking proficiency because the study included neither repeated proficiency assessment nor a comparison condition. The contribution is a contextual account of how audiovisual resources become material for interaction. Instruction should connect manageable video segments to purposeful speaking, equitable participation, and feedback, while subsequent research evaluates learning over time.

Keywords: YouTube Learning Media; teacher mediation; learner participation; EFL speaking

I. INTRODUCTION

English language teaching in Bangladesh continues to raise questions about the relationship between communicative curricular aspirations and classroom opportunities to use English. Hamid and Baldauf (2008) examined barriers to communicative language teaching, while Rahman et al. (2019) discussed interconnected challenges involving policy, curriculum, assessment, and teacher preparation. These accounts provide a context for investigating classroom practice without implying uniform conditions across institutions. In Foundation English, the immediate pedagogical concern is how learners can use available language to formulate ideas, listen to others, and sustain exchanges. Resources matter insofar as teachers and learners can turn them into opportunities for such activity.

For students entering university, a speaking activity may require several kinds of work simultaneously: understanding a topic, finding relevant vocabulary, organizing a response, and managing interaction with classmates. A shared audiovisual text can supply a topic, a communicative situation, and examples of spoken language. It can also introduce unfamiliar

accents, rapid delivery, or cultural assumptions that make participation more difficult. The educational question therefore extends beyond whether students like watching videos. It concerns what happens between exposure to a resource and an opportunity to speak, including the assistance needed by learners who cannot immediately use the material.

YouTube provides teachers with access to conversations, interviews, speeches, documentaries, and other audiovisual genres that can be connected to classroom tasks. Watkins and Wilkins (2011) describe applications of YouTube in EFL instruction, and Sydorenko et al. (2024) examine teachers' perspectives on audiovisual input more broadly. Both direct attention to instructional decisions surrounding the resource. Choosing an appealing clip is only one decision among many. Teachers also determine the focus of viewing, decide when explanation is necessary, and organize subsequent interaction. These decisions are particularly relevant when the objective is speaking rather than comprehension alone.

The distinction between access and use also has implications for evaluating technology in language education. A learner may understand a video without contributing to discussion, or contribute enthusiastically while continuing to make grammatical errors. Similarly, a favorable questionnaire response can indicate perceived usefulness without establishing improvement on a speaking assessment. Hiver et al. (2024) identify substantial variation in the definitions and methods used in language engagement research. Clear distinctions among perception, participation, and learning are consequently necessary if evidence from classroom technology use is to support appropriately bounded conclusions.

Research on perceptions of YouTube is already available in Bangladesh and other EFL contexts. Tahmina (2023) investigated English learners' perceptions at Jagannath University, while Sakkir et al. (2020), Ariyanti and Apoko (2022), and Hudri et al. (2025) examined favorable and critical views in other settings. The existence of this work rules out a general claim that learners' opinions about YouTube remain unexplored. A further perception study becomes more informative when it explains how classroom activity is organized and where positive views coexist with difficulties. The present investigation therefore gives particular attention to teacher decisions and the distribution of speaking opportunities.

The study brings together questionnaire accounts from students, interviews with faculty members, and observations of Foundation English lessons. These sources address different aspects of the classroom experience. Student accounts identify perceived support and difficulty; faculty accounts explain instructional intentions and concerns; observations describe particular instances of participation and mediation. Their combination allows agreement and tension to be examined, although it does not remove the limitations of each source. The analysis is descriptive and exploratory. It does not estimate the independent effect of YouTube or compare it with another instructional treatment.

The contribution is a contextual explanation of the work through which selected videos became resources for speaking. That explanation includes unsuccessful or uneven aspects of participation, because a lively classroom can still leave some learners with few opportunities to speak. It also recognizes that a video, teacher guidance, peer interaction, and feedback occur together in actual lessons. Studying this configuration can inform instructional decisions even when its separate components cannot be causally isolated. Three research questions guide the report:

1. How do teachers integrate YouTube videos into speaking activities in Foundation English classes?
2. What benefits do students and faculty perceive, and what patterns of speaking participation are described in classroom observations?
3. What instructional and technological constraints shape the use of YouTube for speaking practice?

II. THEORETICAL FRAMEWORK

Audiovisual resources and mediated activity

The framework treats audiovisual material as a resource within socially organized activity. Vygotsky's (1978) account of mediation directs attention to the relationship among learners, other participants, and the cultural tools through which activity is conducted. Applied here, the analytical focus is the use teachers and students make of a video within a speaking lesson. The recording provides language and contextual information, while classroom participants interpret that material and respond to one another. The unit of interest is therefore the instructional activity surrounding viewing, rather than the platform considered independently of its use.

This distinction is important when the term scaffolding is applied to educational technology. Wood et al. (1976) describe how tutoring can assist learners in managing a task that would otherwise be difficult. Van de Pol et al. (2010) identify contingency, fading, and transfer of responsibility as central features of scaffolding. A prerecorded speaker does not adjust an explanation in response to the comprehension of a particular learner. Responsive assistance occurs when a teacher or peer notices a difficulty and provides relevant help. Pausing, clarifying, and prompting may therefore be examined as mediation, but the presence of these actions alone does not demonstrate a complete scaffolding process.

The present framework uses this theoretical distinction to avoid attributing agency to the technology. A video can make an expression available for attention, but a teacher may need to explain its meaning, a task must provide a reason to use it, and interlocutors must create space for a response. Such a sequence describes a plausible relationship among resources and activity. It is not a developmental progression established by the current evidence. In particular, the study does not document a systematic withdrawal of support or demonstrate that learners later performed independently in a new setting.

Comprehension support and opportunities for interaction

Mayer's (2021) cognitive theory of multimedia learning emphasizes how learners process words and images under constraints on cognitive capacity. This perspective cautions against assuming that more audiovisual information necessarily produces better learning. For the present study, it motivates attention to selection and manageability: unfamiliar speech, written text, and visual information may compete for attention. The theory is used as a lens for understanding instructional choices, not as a model directly tested through cognitive measures. The study did not assess cognitive load or compare alternative multimedia designs.

Research on captions offers a more specific basis for considering comprehension support. Montero Perez et al. (2013) synthesized evidence concerning captioned video in listening and vocabulary learning, while Montero Perez (2022) reviewed audiovisual input and on-screen

text in second and foreign language learning. This literature supports examining captions as a potentially useful resource, with attention to learning conditions and outcomes. It does not justify treating comprehension assistance as equivalent to speaking development. Reading an expression on screen, recognizing it in speech, and using it appropriately in an exchange are related but distinct demands.

Interaction supplies opportunities that a recording alone cannot provide. Loewen and Sato (2018) review the roles of input, output, negotiation, and feedback in instructed second language acquisition. Their account helps explain why the activity following viewing is analytically important. A learner's response can become the basis for clarification, elaboration, or corrective feedback from another person. In the current framework, a clip supplies shared material for this process. Whether productive interaction follows depends on task organization, comprehension, and access to turns, rather than on audiovisual exposure alone.

Learners' own experiences may also help connect a shared resource with purposeful talk. Lambert et al. (2017) investigated engagement in tasks using learner-generated content and found benefits relative to teacher-generated content in their study. This finding provides a rationale for examining moments when learners connect video topics with their own lives. It is not evidence that every personal response constitutes deeper learning. Here, examples involving familiar places, community leadership, and everyday social expectations are examined as opportunities to formulate meanings beyond the wording of a recording.

Engagement, confidence, and autonomy

Engagement is a multidimensional construct rather than a synonym for enjoyment or visible activity. Philp and Duchesne (2016) distinguish behavioral, cognitive, social, and emotional dimensions in language tasks. Hiver et al. (2024) further emphasize the need for precise conceptualization and operationalization. These distinctions define the scope of the present observations. Taking a turn, asking a question, and responding to a peer provide evidence of participation within an activity. They offer a narrower basis for interpretation than validated measures of engagement across dimensions and over time.

Confidence and willingness to communicate also require careful treatment. MacIntyre et al. (1998) conceptualize willingness to communicate as situated and distinguish it from linguistic competence. This helps explain why students can participate while making errors and why greater ability does not guarantee participation. The present study did not administer a validated willingness-to-communicate scale. Faculty descriptions of greater willingness and student expressions of confidence are therefore reported as perceptions. Observed volunteering illustrates behavior in particular lessons, without establishing a stable change in learners' psychological dispositions.

Time is another important boundary. Aubrey et al. (2022) examined engagement across speaking tasks over a longitudinal period, illustrating the value of studying variation beyond a single activity. Four observations in the current study cannot establish whether interest persisted, diminished, or transferred to other lessons. The same restriction applies to autonomy. Reinders and White (2016) discuss the relationship between technology and learner autonomy in terms that extend beyond access to resources. Requests for additional videos may indicate interest, but do not demonstrate independent goal setting, monitoring, evaluation, or sustained practice.

Research gap and conceptual relationships

The YouTube literature addresses several different evidential questions. Perception studies such as Ariyanti and Apoko (2022) and Hudri et al. (2025) investigate learners' views of speaking-related use. Saed et al. (2021) and Menggo et al. (2025) include assessments of speaking in comparative instructional research. Kaboocha and Elyas (2018) examine vocabulary learning and perceptions in a different instructional context. These studies cannot be pooled conceptually as interchangeable demonstrations of speaking improvement. Their methods and outcome measures determine the conclusions that can be drawn from them.

Recent synthesis also reinforces the need to consider the characteristics of audiovisual instruction. Sutton and Webb (2026) reviewed audiovisual input across several areas of language learning and found that video category moderated outcomes in their analysis. Their synthesis used within-group change, which addresses a different question from the relative effectiveness of matched instructional conditions. Its relevance here is the importance of attending to what learners watch and how outcomes are assessed. It does not supply a missing effect estimate for the present classes or validate claims about unmeasured speaking development.

The gap addressed in this article is consequently specific: an account of how teacher mediation, learner participation, and implementation difficulties relate within the participating Foundation English classes. It is not a claim that YouTube, Bangladesh, or learner perceptions have received no prior attention. The analysis connects teachers' stated purposes with episodes of actual classroom use and with students' reported experiences. This allows favorable accounts to be qualified by the assistance required to understand material and by the uneven distribution of speaking opportunities.

Conceptually, selection and preparation make a video relevant to a speaking objective; explanation during viewing helps learners interpret demanding material; subsequent interaction provides opportunities to formulate responses and receive feedback. Proficiency differences, cultural familiarity, access to turns, and technical continuity can constrain each connection. These relationships organize the interpretation rather than constitute tested causal paths. They lead to a practical analytical question: under what observed conditions did video input become material for student talk, and where did that process remain incomplete? The three research questions examine different aspects of this shared problem.

The framework also separates the levels at which evidence is available. Perceived usefulness concerns an individual's account of an experience. Classroom participation concerns actions situated within a task and a particular group. Speaking development concerns change in ability across time and potentially across tasks. Agreement between the first two levels can strengthen an interpretation of classroom experience, but cannot establish the third without appropriate assessment. This separation informs both the source comparison and the wording of the findings. It prevents a general positive evaluation from absorbing distinctions that matter for research design and pedagogical decision making.

III. METHOD***Research design***

The study used an exploratory descriptive design combining a student questionnaire, semi-structured faculty interviews, and non-participant classroom observations. It examined regular Foundation English instruction in which YouTube videos were used for speaking practice. The questionnaire included closed and open-ended items, while interviews and observations supplied descriptions of instructional activity. No treatment allocation, comparison condition, or repeated assessment of speaking proficiency was included. The design therefore supports an account of practices, experiences, and constraints within the participating setting.

We bring the sources together through an explicit comparison of their reported findings in Table 2. Fetters et al. (2013) explain how joint displays can make relationships among data sources visible. Here, the display is a synthesis at the reporting stage: it distinguishes agreement, complementary information, and interpretive limits. The sequencing and priority of the original data collection are not sufficiently documented to establish a convergent or sequential mixed-methods subtype. The descriptive designation reflects that limitation and the scope of the available evidence.

Participants

The student sample comprised 103 first-year undergraduates associated with Foundation English provision at a private university in Bangladesh. They came from six departments, identified in the study records as English, Economics, CSE, Pharmacy, DBA, and Mathematics. Five English language teaching faculty members also participated; they had experience teaching foundation courses and familiarity with YouTube. Four classes taught by participating instructors were observed. The student cohort, faculty interview sample, and observed lessons represent different sources of evidence and should not be treated as equivalent analytical units.

The recruitment procedure, departmental distribution, and number of students invited were not recorded in sufficient detail to evaluate response coverage. The exact stage of students' course participation and the institutional allocation of the faculty sample also remain unresolved in the documentation. Baseline proficiency and prior YouTube experience are unavailable for subgroup comparison. These limitations restrict interpretation to the participating context. Following the concerns about sampling and generalizability raised by Andringa and Godfroid (2020), the findings are not presented as representative of Bangladeshi university students or institutions.

Table 1-Data sources and scope of evidence

Source	Reported sample or record	Analytical use	Principal limitation
Student questionnaire	103 first-year undergraduates	Perceptions of use, support, and difficulties	Item denominators and complete response distributions unverified
Faculty interviews	5 faculty members; written notes; 20–30 minutes each	Instructional choices, perceived participation, and implementation needs	Institutional allocation and detailed analytic procedure unresolved
Classroom observations	4 lessons; approximately 60	Teacher mediation, speaking activity, and classroom constraints	Limited coverage; no standardized proficiency assessment

	minutes each; field notes		
Instructional materials	16 videos introduced across 12 speaking classes	Context for the observed activities	Complete video inventory and exposure by participant not documented

Instruments

The questionnaire elicited views about YouTube-supported speaking through closed and open-ended questions. Reported topics included confidence, comprehension, pronunciation, vocabulary, fluency, grammar, subtitles, course integration, and learning difficulties. The aggregate summaries used response labels including Always, Usually, and Sometimes. The full wording, complete response options, item count, and instrument-development record were not available for verification. Accordingly, the questionnaire is treated as a source of descriptive perceptions, not a validated measure of speaking proficiency, motivation, or engagement. No composite score or reliability coefficient is asserted.

The faculty interview guide addressed classroom uses of videos, content selection, perceived participation, difficulties, curricular support, and implementation needs. Interviews were documented in written notes. Classroom observation focused on lesson organization, teacher support, speaking activities, and difficulties encountered during use. Lesson plans and selected videos provided contextual material and were reviewed before observation. Complete versions of the interview guide and observation schedule are needed for replication; their absence also limits assessment of how consistently topics and behaviors were elicited across participants and lessons.

Data collection methods

Sixteen YouTube videos were introduced across twelve speaking classes during an academic semester. Teachers and researchers selected material collaboratively with attention to syllabus relevance, language level, communicative objectives, authenticity, length, and potential for discussion. The repertoire included conversations, interviews, travel material, speeches, and cultural topics. Source videos lasted approximately 3–40 minutes. Two or three short videos could be combined, while selected segments were used from longer recordings. These lengths describe source material, not the actual viewing time or exposure received by individual students.

Instruction involved preparation before viewing, support during viewing, and subsequent speaking activity. Preparation included stating objectives, activating background knowledge, and introducing vocabulary or pronunciation features. During viewing, teachers could pause the video, check comprehension, and explain unfamiliar expressions or cultural references. Follow-up activities included pair and group discussions, role-play, interviews, retelling, presentations, debates, and problem solving. This describes the reported instructional repertoire; it does not imply that every activity occurred in every lesson or that a standardized treatment was delivered uniformly.

Four regular classes were observed for approximately sixty minutes each. The researcher did not participate in the classroom activities and completed detailed field notes immediately after each lesson. These observations covered only part of the twelve-class instructional sequence. Calendar dates, selection of the observed lessons, participant overlap, and individual viewing exposure are not fully documented. The available evidence consequently

supports description of particular episodes rather than comprehensive monitoring of every class or learner across the semester.

Students received an explanation of the research purpose and instructions for completing the questionnaire. Some open-ended responses were written in Bengali and translated into English by the researchers. Individual faculty interviews lasted approximately 20–30 minutes, with permission obtained for written note-taking. Audio recording and verbatim transcription were not reported. The timing of the questionnaire and interviews relative to the observed lessons, and the procedure for checking translated meanings, are not specified in the available records. This limits fine-grained temporal and linguistic interpretation.

Data analysis techniques

Closed-item responses were summarized descriptively. Figures 1–3 retain the reported percentages and item labels. An arithmetic check identified inconsistent category totals in Figure 2; item counts, valid denominators, and missing-response records were unavailable for reconciliation. Its bars therefore display the supplied values on an absolute percentage scale, without normalization to 100%. Figures 1 and 3 retain their separate item entries. The figures document the reported summaries, with their limitations stated in the notes. No counts were reconstructed, categories imputed, or inferential tests added; 103 is the sample size, not a verified denominator for every item.

The qualitative material was originally interpreted thematically. In this report, the existing findings are organized under the three research questions, with questionnaire accounts, faculty perspectives, and observation episodes distinguished throughout. The underlying coding orientation, researcher roles, and theme-development procedures are incompletely documented. Braun and Clarke (2021) emphasize coherence between the form of thematic analysis adopted and its reporting. We therefore describe the available interpretation without claiming an undocumented coding protocol, independent coding, saturation assessment, or retrospective validation procedure.

Table 2 compares the reported evidence across sources and retains difficulties alongside favorable accounts. It is not a fresh analysis of raw response files or transcripts. Statements are paraphrased at source-type level because participant identifiers and the verbatim status of quotations were not established. Material attributed only to unspecified student interviews is excluded: the documented student instrument was a questionnaire. These reporting decisions prioritize a traceable distinction between evidence and interpretation, consistent with the transparency concerns articulated by O'Brien et al. (2014).

Ethical considerations

Permission for classroom observation was obtained from the department and participating instructors. Students and faculty were informed about the research purpose, and faculty permitted written interview notes. The available documentation does not establish institutional ethics approval or exemption, voluntary participation and withdrawal arrangements, management of teacher–student power relationships, or confidentiality and retention procedures. These are unresolved reporting requirements rather than assumed safeguards. The present account does not equate access permission with informed consent or claim an approval reference unsupported by the study record.

IV. RESULTS AND DISCUSSION

Results

Teacher integration of YouTube into speaking activities

The observation accounts describe a sequence connecting preparation, viewing, and speaking. Teachers introduced the purpose of the activity and prepared learners for relevant vocabulary and pronunciation features. They directed attention to ideas and expressions that could be used subsequently, rather than leaving the video disconnected from the lesson. During viewing, pauses provided opportunities to explain difficult language, check meaning, and draw attention to communicative features such as intonation and gestures. The input thus became part of a guided activity with an identifiable speaking purpose.

Post-viewing work included pair interaction, group discussion, and presentations. In an activity focused on polite disagreement, students reused expressions from a video in their exchanges. Two students attempted to reproduce the speaker's intonation. Other observed exchanges included follow-up questions, reasons for opinions, and summaries of ideas. These examples show learners using material available in the lesson while responding to an interlocutor. They establish immediate connections between the selected input and task performance, but the records do not show whether the same expressions were retained or used independently later.

Faculty interviews described videos as sources of discussion topics, pronunciation models, and preparation for presentations. Their accounts indicated variation in frequency of use: some instructors described regular integration, whereas others used video occasionally. The common feature was an instructional purpose assigned to the resource, rather than a single uniform method. Teachers also described a need to consider learners' language resources and the suitability of topics. The video repertoire and subsequent tasks therefore reflected decisions about what could be made useful in a particular class.

Faculty identified limited syllabus guidance for selecting digital material and organizing related speaking work. They called for examples of suitable activities, professional development, and closer connections among speaking objectives, teaching resources, and assessment. In answer to the first research question, integration involved planned selection, preparation, explanation during viewing, and opportunities to speak afterward. The accounts also showed variation across instructors and a need for institutional support. Video availability supplied material for instruction, while its classroom use required additional pedagogical decisions.

Perceived benefits and observed speaking participation

Figure 1 presents the reported responses concerning confidence, subtitles, and course integration. Confidence was associated with Always by 56% and Usually by 36%; corresponding subtitle entries were 64% and 28%. The support for integrating YouTube activities into English courses was 97%. These selected responses indicate a favorable pattern in the original summary, but the items have different purposes and incomplete response documentation. They cannot be combined into a single acceptance score, and the reported support percentage does not establish improvement in speaking. Faculty accounts similarly described perceived willingness to contribute after some activities.

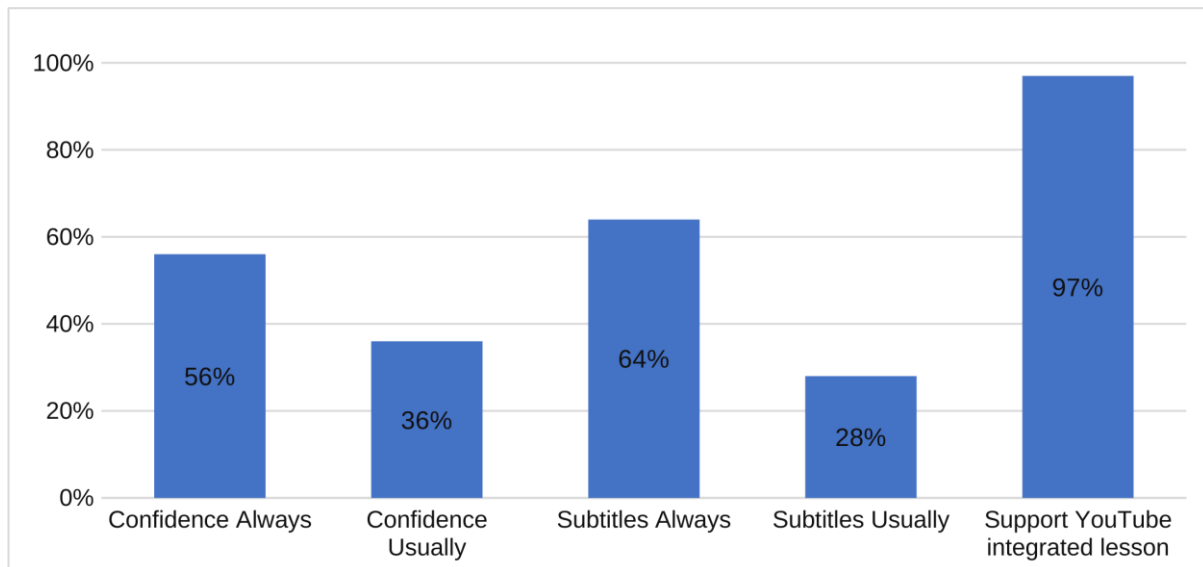


Figure 1-Perceptions toward integration of YouTube in enhancing the Speaking Skill

Figure 2 distinguishes perceived support across seven speaking-related domains. The original Always entries were 69% for phonemes, stress, and intonation; 67% for pronunciation; 64% for comprehension; 56% for confidence; 54% for fluency; 48% for vocabulary; and 36% for grammar. The Always, Usually, and Sometimes values remain visible. However, the category totals range from 83% to 104%, so the chart cannot establish valid response distributions or a comparative hierarchy of learning gains. It identifies the domains addressed by the questionnaire while keeping the numerical discrepancies open to scrutiny.

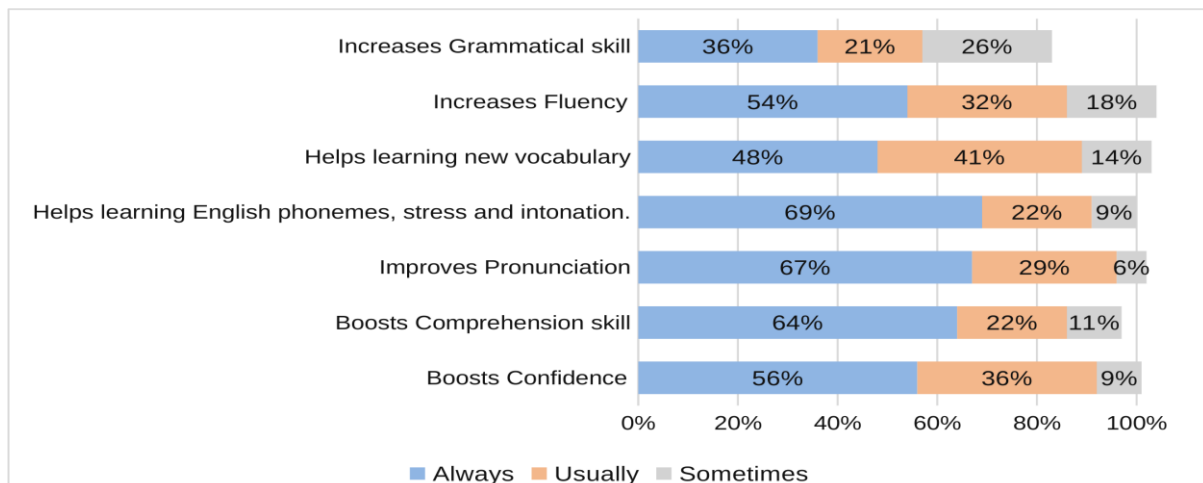


Figure 2-Perceptions Regarding Usage of YouTube in Developing Different Segments

Observation accounts recorded voluntary answers, questions, exchanges among peers, and continued discussion after viewing. A lesson using street interviews prompted students to compare expectations about speaking with strangers across social contexts. Learners continued exchanging opinions while making grammatical errors. The episode is informative because willingness to contribute and linguistic accuracy did not coincide automatically. The

observed outcome was participation in meaning-focused discussion, with students using the shared topic to express views and respond to one another.

A documentary about Bhutan generated discussion of travel preferences and comparisons with Bangladesh. The material supplied a common referent, while learners drew on familiar experiences to develop responses. In another lesson, a leadership talk by Simon Sinek prompted discussion of leaders in students' university and community. Participants supported opinions with personal examples. These episodes included responses that extended beyond repetition of the recording. They show the use of video content as a starting point for locally relevant discussion, without establishing a general increase in fluency.

Students also requested related videos, and faculty described interest in resources for further viewing. Such requests indicate an intention to continue engaging with material. Independent use was not followed systematically, so the accounts do not establish whether students subsequently watched the resources, practised speaking, or monitored their own progress. The second research question is therefore answered through perceived usefulness and documented classroom participation. The evidence concerns opportunities and responses within instruction, while sustained autonomy and independently assessed speaking development remain outside its demonstrated scope.

Instructional and technological constraints

Figure 3 presents twelve reported implementation challenges. Poor network connection had the largest reported entry, at 78%, followed by lack of feedback at 69%, fast video pace at 68%, and loss of concentration at 67%. The figure also retains difficulties involving interaction, vocabulary, subtitles, content suitability, advertisements, authenticity, practicality, and accents. These entries describe the summary; their ordering does not establish reliable differences or population prevalence. Faculty accounts placed these difficulties within the task: learners needed assistance when they could not follow the input or formulate a response.

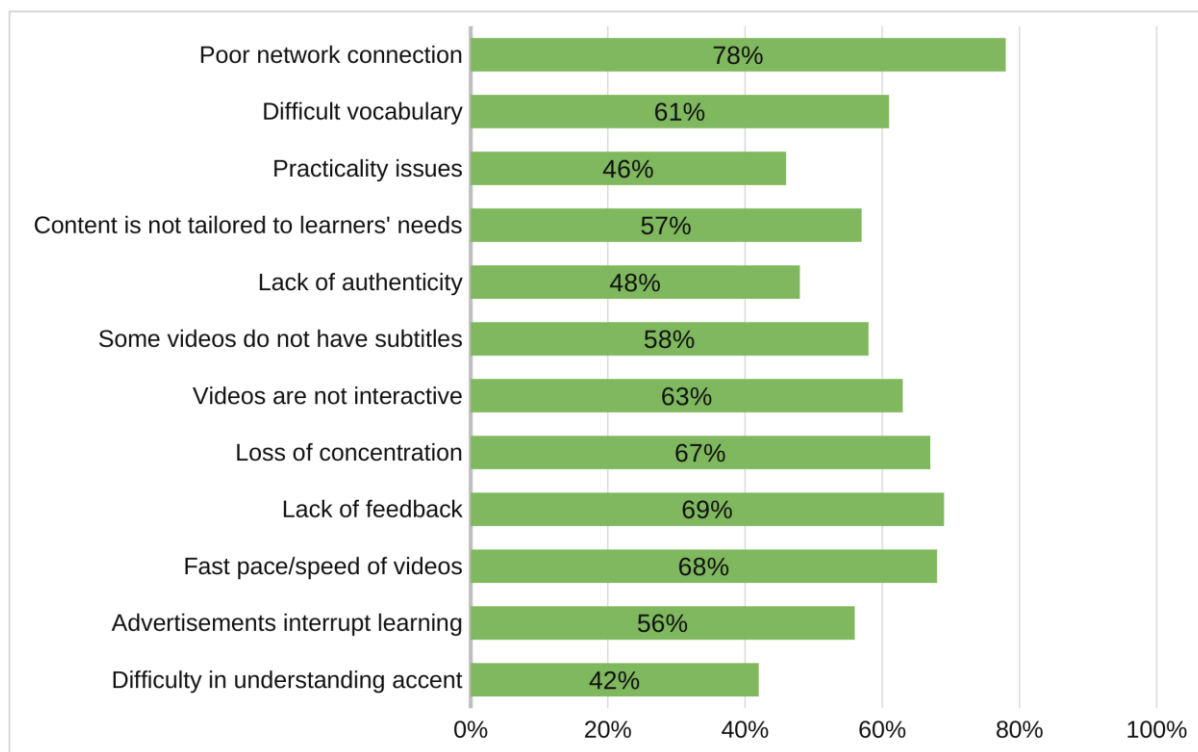


Figure 3-Challenges emerged from integrating YouTube videos

The observation accounts show specific instructional responses to comprehension difficulties. In one lesson, the teacher enabled English subtitles after learners struggled with the pace and accent of the clip. In the Bhutan lesson, Gross National Happiness required explanation because students initially understood the concept too narrowly. These examples demonstrate that visual context did not eliminate linguistic or cultural difficulty. Teacher intervention was needed to restore a usable understanding of the material before discussion could develop meaningfully.

Participation was also uneven within group work. More proficient learners dominated some exchanges, leaving less proficient classmates fewer opportunities to speak. A group could therefore appear active while contributions were concentrated among particular members. This finding limits any interpretation of classroom liveliness as evidence that all learners benefited equally. The records identify an instructional problem concerning the allocation of turns, although they do not provide turn counts or proficiency measurements that would permit a systematic comparison of participation patterns.

Technical interruptions further affected continuity. Internet disruptions interrupted classroom activity, while student accounts identified advertisements and loss of concentration as difficulties. Faculty emphasized that a prerecorded video could neither answer an individual question nor supply contingent corrective feedback. The third research question is thus answered by interacting linguistic, cultural, organizational, and technological constraints. Table 2 brings these findings together with favorable accounts, showing that perceived value and barriers to participation were present within the same instructional setting.

Table 2- Cross-source comparison of reported findings

Focus	Survey summary and figures	Faculty accounts	Observation accounts	Interpretive limit
Confidence	Confidence: Always 56%, Usually 36% (Figure 1); unverified denominators	Greater willingness to contribute described after some activities	Voluntary answers, questions, and peer exchanges	Participation does not establish a proficiency gain
Input support	Subtitles: Always 64%, Usually 28% (Figure 1); difficult input also reported	Need to select suitable language and topics	Subtitles and explanation used when comprehension was difficult	Support was needed; video access alone was insufficient
Using language in tasks	Seven perceived domains (Figure 2); category totals 83%–104%	Videos used as models and discussion prompts	Expressions and intonation reused in pair work	Immediate use does not establish retention or transfer
Continued practice	No verified frequency of independent practice	Requests for related videos described	Learners asked for additional material	Interest does not establish sustained autonomy
Barriers	Twelve reported challenges (Figure 3); network connection 78%	Preparation and responsive support considered necessary	Connectivity interruption and unequal speaking turns	Favorable views coexist with barriers to participation

DISCUSSION

Teacher mediation and the conversion of input into interaction

The central interpretation is that the classroom value of the videos depended on the activity organized around them. Selection created a shared topic and made particular language available; explanation addressed difficulties; subsequent tasks provided occasions to respond. The observed episodes illustrate links among these actions rather than an autonomous instructional effect of the platform. This contribution is useful because a description of positive attitudes alone would not explain how learners moved from watching a recording to exchanging ideas with classmates.

The interpretation is compatible with a sociocultural emphasis on mediated activity (Vygotsky, 1978). It also requires a boundary between responsive assistance and a demonstrated developmental process. Teachers' explanations and pauses are evidence of support within the observed lessons. They do not establish systematic fading or transfer of responsibility as discussed by van de Pol et al. (2010). The most defensible contribution is

therefore an account of mediation during particular activities, including moments when understanding had to be repaired before learners could continue.

This process perspective helps connect the study with the teacher accounts examined by Sydorenko et al. (2024). Selecting an appropriate resource and designing the surrounding task are part of the pedagogical work, not incidental preparation. In the present setting, cultural explanation and language support were especially visible when a clip proved demanding. The Bhutan episode shows why relevance cannot be inferred from a topic alone: learners could discuss travel preferences, yet a central concept still needed explanation. Suitability emerged through the relationship between material, learners, and instructional support.

Participation as a meaningful but bounded outcome

The observation episodes add detail to the favorable perceptions reported in earlier work, including Sakkir et al. (2020) and Tahmina (2023). They show questions, peer responses, personal examples, and reuse of expressions. These actions matter pedagogically because they constitute opportunities to use English within the lesson. Their value does not require presenting them as proof of general proficiency improvement. A bounded account of participation can be informative while acknowledging that accuracy, fluency, and retention were not independently assessed.

The contrast between active discussion and continuing grammatical errors is particularly instructive. It shows why classroom participation and language achievement should be analyzed separately. Students may contribute despite limited accuracy, while technically accurate learners may choose not to speak. This interpretation is consistent with MacIntyre et al.'s (1998) distinction between willingness and competence. It does not establish that the video reduced anxiety or changed willingness as a measured construct. The present evidence concerns what participants described and what the observer recorded in specific instructional circumstances.

The findings also highlight the distribution of opportunities within a task. More proficient students could use the shared topic quickly and dominate discussion, while others contributed less. A group-level impression of enthusiasm can conceal this imbalance. The multidimensional perspective of Philp and Duchesne (2016) encourages attention to how participation is enacted rather than treating visible energy as a comprehensive engagement score. Here, the contrast between lively exchanges and unequal access to turns suggests that evaluation of a video activity should consider who speaks, with whom, and under what support.

Feedback connects participation with a further instructional opportunity. The inability of a recording to respond to an individual learner was identified by faculty as a limitation. In a classroom, that limitation can be addressed through interaction with teachers and peers, although the form and quality of the response require separate attention. Loewen and Sato (2018) provide a basis for considering feedback within language interaction. The present evidence does not establish which feedback moves occurred consistently or which learners used them. It nevertheless identifies why speaking after viewing needs an interlocutor who responds, rather than only an instruction to reproduce language from the screen.

Comprehension, personal relevance, and alternative explanations

Subtitles and explanation appeared as responses to difficulty rather than universal solutions. The caption literature reviewed by Montero Perez (2022) provides a rationale for

investigating comprehension support, but the present observations do not compare subtitle conditions. The teacher's decision to enable subtitles identifies a perceived instructional need and a response to it. Without an independent comprehension measure, it cannot quantify the benefit of that response. Similarly, a cultural explanation may help learners join a discussion while leaving other lexical or pronunciation demands unresolved.

The use of personal examples offers another plausible interpretation of participation. The leadership and travel activities allowed learners to connect shared material with familiar people and places. Lambert et al. (2017) provide relevant evidence that learner-generated content can support task engagement under their study conditions. Our observations suggest a related possibility: a video may be useful partly because it opens a topic to which learners can contribute their own knowledge. This remains an interpretation of classroom episodes, rather than evidence isolating personal relevance as the mechanism responsible for participation.

Several alternative explanations must remain open. Interest may have arisen from an unfamiliar activity, a compelling topic, teacher enthusiasm, or the opportunity to work with peers. Students with stronger language resources may have been more able to use the input and thus more visible in the observations. Teacher support, video content, and discussion were delivered together, preventing their contributions from being separated. These alternatives do not negate the reported experiences; they define what additional evidence would be needed before attributing an outcome specifically to YouTube.

Relationship to learning and autonomy research

Studies by Saed et al. (2021) and Menggo et al. (2025) address speaking outcomes through comparative instructional designs. Their findings establish a relevant research context, but cannot be imported as outcome evidence for these students. Likewise, vocabulary results reported by Kaboocha and Elyas (2018) concern a different measured domain. The present investigation contributes descriptions of implementation and participation that can help interpret how such interventions are enacted. It leaves the magnitude, durability, and transfer of speaking development unresolved.

Sutton and Webb's (2026) synthesis further suggests that audiovisual research should distinguish types of material and learning outcomes. The present repertoire included interviews, documentary content, and a public talk, but exposure was not documented sufficiently to compare genres. The evidence therefore cannot identify a superior type of video or an optimal viewing duration. The practical implication is to align the resource with the intended activity and document that alignment, rather than assume that all videos constitute a comparable instructional treatment.

Requests for additional videos should similarly be interpreted as interest in further activity. Reinders and White (2016) locate autonomy in learners' management of learning, not merely in access to technology. The study does not show how students selected materials or evaluated their understanding outside class. Shadiev et al. (2020) review mobile-assisted language learning in familiar, authentic environments, but such environments extend beyond the classroom episodes documented here. Evidence of learning across settings would require observation or records of actual independent practice, not inference from a request for resources.

Instructional planning can begin with the speaking purpose rather than the availability of an attractive clip. A teacher can identify the response learners should produce, select a manageable segment, and prepare the language or background knowledge needed to enter the task. The observed difficulty with pace, accent, and cultural meaning supports attention to these demands before and during viewing. This is a design implication from the reported cases, not evidence that a fixed sequence or particular clip length is optimal for every class.

During viewing, support can be responsive to the difficulty that emerges. A brief pause may be useful when a phrase blocks understanding; a cultural explanation may be needed when the topic itself is misunderstood. Captions should be considered in relation to the lesson objective and learner needs rather than assumed to remove every comprehension problem. Mayer's (2021) account of multimedia learning provides a theoretical reason to consider the combined demands of speech, images, and text. The practical aim is to preserve enough comprehension for subsequent speaking without allowing explanation to consume the entire activity.

After viewing, task organization should make participation accessible. Planned turns, rotating roles, brief preparation time, or prompts requiring a follow-up question are possible responses to the uneven participation observed here. These specific arrangements were not evaluated in the study and should be treated as proposals for practice. Their purpose is to make opportunities visible and distribute them more deliberately. Teachers can also distinguish feedback on ideas and interaction from feedback on language form, so that an enjoyable exchange does not substitute for attention to the speaking objective.

Institutional support can address both preparation and continuity. Shared examples of video selections and related tasks would respond to faculty requests for guidance, while professional learning could focus on matching resources to proficiency and communicative aims. Reliable access and a feasible alternative activity would help teachers manage interruptions. Evaluation should attend to participation and task performance as well as student satisfaction. These recommendations concern the conditions needed to implement instruction; they are not a claim that access to the platform alone produces equitable or durable learning benefits.

The study draws on students from one private university, five faculty interviews, and four observed lessons. Recruitment, participant overlap, faculty-site allocation, and course timing were incompletely documented. These limitations constrain judgments about whose experiences the findings represent. As Andringa and Godfroid (2020) emphasize, sampling choices shape the reach of applied linguistic knowledge. The current account can inform questions for comparable Foundation English settings, but cannot support a population estimate or a general claim about university instruction throughout Bangladesh.

The numerical and qualitative records impose further restrictions. Figures 1–3 preserve the reported survey values, but unresolved denominators and inconsistent category totals limit their interpretation. They cannot support validated prevalence estimates or comparisons of learning outcomes. Incomplete instrument documentation, coding records, translation checks, and source attribution reduce reproducibility. Researcher involvement in selecting materials may have influenced the activities and their interpretation; the observer's presence and participants' expectations may also have shaped responses. The reporting

limitations require verified source records and transparent accounts of the procedures actually undertaken.

Future work should link instruments, participants, lessons, and source records explicitly, while documenting consent and ethical oversight. A complete video inventory could record selected segments, actual viewing time, caption use, tasks, and interruptions. Repeated observations could then examine how speaking opportunities are distributed and whether interest persists across lessons, following the longitudinal concern illustrated by Aubrey et al. (2022). Independently rated speaking assessments and suitable comparison conditions would address development more directly. Follow-up records of independent practice could test whether requests for further videos become sustained learning activity.

A comparison study would also need to specify what differs between conditions. If one group receives more discussion time, feedback, or preparation, a difference in performance cannot be assigned solely to the video resource. Matching these opportunities and documenting implementation would make the comparison more interpretable. Conversely, a process study could examine variation among learners within the same activity through repeated records of turns and responses. These approaches answer complementary questions about outcomes and enactment. Selecting the design after defining the intended claim would help future work preserve the distinction between an appealing instructional experience and demonstrable development in speaking.

Future observation should also attend to learners who speak less. Silence can reflect difficulty understanding the input, uncertainty about a response, limited access to a turn, or a deliberate choice to listen. The present notes cannot distinguish these possibilities. Combining observations with brief accounts from those learners would help identify the support needed, while avoiding an assumption that visible talk fully represents every learner's experience of the speaking activity.

V. CONCLUSION

This study examined how YouTube was incorporated into Foundation English speaking classes through teacher preparation, explanation during viewing, and subsequent interaction. Student and faculty accounts described perceived usefulness, while observation episodes documented questions, peer exchanges, personal examples, and immediate use of language from videos. The same evidence identified demanding input, unfamiliar cultural references, technical interruptions, limited responsive feedback, and unequal speaking opportunities. The findings therefore support a contextual interpretation of YouTube as a supplementary resource whose classroom value depends on how teachers organize and support participation. They do not establish causal gains in speaking proficiency, sustained autonomy, or effectiveness across Bangladeshi higher education. Unreconciled survey distributions and incomplete methodological and ethical documentation further limit the strength and reproducibility of the account. The contribution lies in identifying the instructional work that connects audiovisual input with student talk and the conditions that can interrupt that connection. For practice, purposeful tasks, manageable input, equitable access to speaking, and feedback deserve attention alongside resource selection. For research, verified source records, repeated observation, independently assessed speaking outcomes, and follow-up

across settings are needed to determine whether these classroom opportunities contribute to durable learning.

VI. REFERENCES

- Andringa, S., & Godfroid, A. (2020). Sampling bias and the problem of generalizability in applied linguistics. *Annual Review of Applied Linguistics*, 40, 134–142. <https://doi.org/10.1017/S0267190520000033>
- Ariyanti, P., & Apoko, T. W. (2022). The application of YouTube media in learning speaking: Students' perceptions. *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, 10(2), 1324–1330. <https://doi.org/10.24256/ideas.v10i2.3279>
- Aubrey, S., King, J., & Almukhailid, H. (2022). Language learner engagement during speaking tasks: A longitudinal study. *RELC Journal*, 53(3), 519–533. <https://doi.org/10.1177/0033688220945418>
- Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qualitative Research in Psychology*, 18(3), 328–352. <https://doi.org/10.1080/14780887.2020.1769238>
- Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs—Principles and practices. *Health Services Research*, 48(6 Pt 2), 2134–2156. <https://doi.org/10.1111/1475-6773.12117>
- Hamid, M. O., & Baldauf, R. B., Jr. (2008). Will CLT bail out the bogged down ELT in Bangladesh? *English Today*, 24(3), 16–24. <https://doi.org/10.1017/S0266078408000254>
- Hiver, P., Al-Hoorie, A. H., Vitta, J. P., & Wu, J. (2024). Engagement in language learning: A systematic review of 20 years of research methods and definitions. *Language Teaching Research*, 28(1), 201–230. <https://doi.org/10.1177/13621688211001289>
- Hudri, M., Iman, N. H., & Ilham, I. (2025). Students' perceptions of YouTube as a learning medium for enhancing speaking skills in English language acquisition at Universitas Muhammadiyah Mataram. *IJLECR (International Journal of Language Education and Cultural Review)*, 11(1), 79–88. <https://doi.org/10.21009/ijlecr.v11i1.53584>
- Kabooha, R., & Elyas, T. (2018). The effects of YouTube in multimedia instruction for vocabulary learning: Perceptions of EFL students and teachers. *English Language Teaching*, 11(2), 72–81. <https://doi.org/10.5539/elt.v11n2p72>
- Lambert, C., Philp, J., & Nakamura, S. (2017). Learner-generated content and engagement in second language task performance. *Language Teaching Research*, 21(6), 665–680. <https://doi.org/10.1177/1362168816683559>
- Loewen, S., & Sato, M. (2018). Interaction and instructed second language acquisition. *Language Teaching*, 51(3), 285–329. <https://doi.org/10.1017/S0261444818000125>

- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562. <https://doi.org/10.1111/j.1540-4781.1998.tb05543.x>
- Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
- Menggo, S., Widiastri, D. A., Krismayani, N. W., & Susanto, I. (2025). Exploring YouTube-based learning in boosting EFL learners' speaking competence and learning autonomy. *Theory and Practice in Language Studies*, 15(2), 392–402. <https://doi.org/10.17507/tpls.1502.09>
- Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching*, 55(2), 163–192. <https://doi.org/10.1017/S0261444821000501>
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720–739. <https://doi.org/10.1016/j.system.2013.07.013>
- O'Brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A., & Cook, D. A. (2014). Standards for reporting qualitative research: A synthesis of recommendations. *Academic Medicine*, 89(9), 1245–1251. <https://doi.org/10.1097/ACM.0000000000000388>
- Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks in the language classroom. *Annual Review of Applied Linguistics*, 36, 50–72. <https://doi.org/10.1017/S0267190515000094>
- Rahman, M. M., Islam, M. S., Karim, A., Chowdhury, T. A., Rahman, M. M., Seraj, P. M. I., & Singh, M. K. M. (2019). English language teaching in Bangladesh today: Issues, outcomes and implications. *Language Testing in Asia*, 9, Article 9. <https://doi.org/10.1186/s40468-019-0085-8>
- Reinders, H., & White, C. (2016). 20 years of autonomy and technology: How far have we come and where to next? *Language Learning & Technology*, 20(2), 143–154. <https://doi.org/10.64152/10125/44466>
- 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), Article e07543. <https://doi.org/10.1016/j.heliyon.2021.e07543>
- Sakkir, G., Dollah, S., & Ahmad, J. (2020). Students' perceptions toward using YouTube in EFL classrooms. *Journal of Applied Science, Engineering, Technology, and Education*, 2(1), 1–10. <https://doi.org/10.35877/454RI.asci2125>
- Shadiev, R., Liu, T., & Hwang, W.-Y. (2020). Review of research on mobile-assisted language learning in familiar, authentic environments. *British Journal of Educational Technology*, 51(3), 709–720. <https://doi.org/10.1111/bjet.12839>

- Sutton, D., & Webb, S. (2026). The effects of audiovisual input on second language learning: A meta-analysis. *Studies in Second Language Acquisition*, 48(2), 571–595. <https://doi.org/10.1017/S0272263126101612>
- Sydorenko, T., Cárdenas-Claros, M. S., Huntley, E., & Montero Perez, M. (2024). Audiovisual input in language learning: Teachers' perspectives. *Language Learning & Technology*, 28(1), 1–25. <https://doi.org/10.64152/10125/73587>
- Tahmina, T. (2023). Students' perception of the use of YouTube in English language learning. *JOLLT Journal of Languages and Language Teaching*, 11(1), 151–159. <https://doi.org/10.33394/jollt.v11i1.6883>
- van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review*, 22, 271–296. <https://doi.org/10.1007/s10648-010-9127-6>
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.
- Watkins, J., & Wilkins, M. (2011). Using YouTube in the EFL classroom. *Language Education in Asia*, 2(1), 113–119. https://doi.org/10.5746/LEiA/11/V2/I1/A09/Watkins_Wilkins
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>



© 2018 by the journal. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC-BY SA) license (<http://creativecommons.org/licenses/by/4.0/>).