International Journal of Education, Social Science & Humanities.
Finland Academic Research Science Publishers
ISSN: 2945-4492 (online) | (SJIF) = 8.09 Impact factor
Volume-14| Issue-11| 2026 Published: |22-11-2026|

INTEGRATING AUTHENTIC VIDEO MATERIALS INTO ESL INSTRUCTION: DEVELOPING LANGUAGE PROFICIENCY AND INTERCULTURAL AWARENESS

Authors: **Uksinbaeva Malika Imalatdinovna**
Arabbayeva Xusnida Arabbayevna
English Teachers, Academic Lyceum of the University of World Economy and Diplomacy

Abstract: Background. Authentic video (film, television, news, documentaries, and online clips produced for native speakers) is widely used in English as a second language (ESL) classrooms, yet its pedagogical value depends on how it is selected, scaffolded, and assessed. **Purpose.** This structured narrative review synthesizes theoretical and empirical research on authentic video in ESL instruction, asking (RQ1) how it supports listening comprehension, vocabulary, and oral proficiency; (RQ2) how it develops intercultural awareness; and (RQ3) which design conditions moderate these outcomes. **Method.** Foundational theoretical works and empirical studies and meta-analyses on video, captioning, and intercultural competence were identified through targeted searches, screened against explicit inclusion criteria, and analyzed thematically. **Results.** Four themes emerged: (a) multimodal input aids comprehension and incidental vocabulary learning, especially with same-language captions and repeated viewing; (b) video offers rich, contextualized exposure to pragmatic and cultural norms, but intercultural gains appear only when viewing is paired with guided reflection and comparison; (c) learner proficiency, text difficulty, and cognitive load are key moderators; and (d) task design, not the video itself, determines whether viewing becomes active language use. **Conclusion.** A three-phase framework (pre-viewing, while-viewing, post-viewing) integrating language-focused and culture-focused tasks is proposed, with implications for teachers and a research agenda for classroom-based studies.

Keywords: authentic video; ESL instruction; listening comprehension; captioning; intercultural communicative competence; multimodal input; task design.

Аннотация: Актуальность. Аутентичные видеоматериалы (фильмы, телепрограммы, новости, документальные фильмы и онлайн-ролики, созданные для носителей языка) широко применяются на занятиях по

International Journal of Education, Social Science & Humanities.
Finland Academic Research Science Publishers
ISSN: 2945-4492 (online) | (SJIF) = 8.09 Impact factor
Volume-14| Issue-11| 2026 Published: |22-11-2026|

английскому как второму языку (ESL), однако их педагогическая ценность зависит от того, как они отбираются, как организуется опора для учащихся и как оцениваются результаты. **Цель.** В данном структурированном нарративном обзоре обобщаются теоретические и эмпирические исследования об использовании аутентичного видео в обучении ESL. Рассматриваются следующие вопросы: (RQ1) как видео способствует развитию понимания на слух, лексики и устной речи; (RQ2) как оно формирует межкультурную осведомлённость; (RQ3) какие условия организации обучения влияют на эти результаты. **Метод.** Фундаментальные теоретические работы, эмпирические исследования и метаанализы по видео, субтитрам и межкультурной компетенции были целенаправленно отобраны, проверены на соответствие чётким критериям включения и проанализированы тематически. **Результаты.** Выделены четыре темы: (а) мультимодальный input способствует пониманию и неосознанному усвоению лексики, особенно при наличии субтитров на изучаемом языке и повторном просмотре; (б) видео предоставляет богатый контекстуализированный материал о прагматических и культурных нормах, однако межкультурные результаты проявляются только тогда, когда просмотр сопровождается направляемой рефлексией и сопоставлением; (в) ключевыми модераторами являются уровень владения языком, сложность материала и когнитивная нагрузка; (г) именно структура заданий, а не само видео, определяет, станет ли просмотр активным использованием языка. **Выводы.** Предложена трёхэтапная модель (до просмотра, во время просмотра, после просмотра), объединяющая языковые и культурные задания, а также представлены рекомендации для преподавателей и программа дальнейших классных исследований.

Ключевые слова: аутентичное видео; обучение ESL; понимание на слух; субтитры; межкультурная коммуникативная компетенция; мультимодальный input; структура заданий.

INTRODUCTION

Learners of English as a second language rarely hear the language the way native speakers use it. Textbook dialogues are typically slowed, scripted, and stripped of the hesitations, regional accents, overlapping speech, and cultural references that characterize real communication. Authentic video materials, defined here as audiovisual texts created for native-speaker audiences rather than for language teaching (Gilmore, 2007), offer a way to close this gap. Their combination of speech, image, gesture, setting, and (often) written captions gives learners

context for interpreting meaning and exposes them to language in use (Berk, 2009; Canning-Wilson, 2000).

1.1 Theoretical background

Several theoretical perspectives explain why video may benefit second language learning. The input hypothesis holds that acquisition proceeds when learners understand input slightly beyond their current competence (Krashen, 1982); visual context can make such input comprehensible. Dual coding theory proposes that information encoded both verbally and visually is better retained (Paivio, 1986), and the cognitive theory of multimedia learning similarly argues that people learn more deeply from words and pictures together when the design avoids overloading working memory (Mayer, 2009; Sweller, 1988). From an interactionist standpoint, input becomes acquisitional through negotiation of meaning (Long, 1996) and learner output (Swain, 1985), which suggests that video must be paired with speaking and writing tasks rather than watched passively.

Video is also a cultural text. Kramsch (1993) argued that language teaching cannot be separated from the teaching of culture, and Byram (1997) conceptualized intercultural communicative competence as the knowledge, attitudes, and skills needed to interpret and relate to other cultures while mediating between them. The Common European Framework of Reference Companion Volume likewise gives mediation and plurilingual and pluricultural competence a prominent place (Council of Europe, 2020).

1.2 Previous research and gap

Empirical work has examined video chiefly in two strands. The first concerns language outcomes: captioned video and listening comprehension (Hayati & Mohmedi, 2011; Winke et al., 2010), vocabulary gains from viewing (Peters & Webb, 2018; Sydorenko, 2010), and extensive viewing as an out-of-class resource (Webb, 2015). A meta-analysis of captioned video research has summarized much of this evidence (Montero Perez et al., 2013). The second strand concerns culture and film in the classroom (Chapple & Curtis, 2000; Herron & Hanley, 1992). However, these strands are seldom integrated: studies of linguistic gains rarely address intercultural learning, and discussions of culture rarely specify the task conditions that make video effective for language development. Teachers therefore receive few evidence-based guidelines for designing lessons that serve both goals.

1.3 Purpose and research questions

This article synthesizes the literature to address three research questions:

RQ1. How does the use of authentic video affect ESL learners' listening comprehension, vocabulary, and oral proficiency?

International Journal of Education, Social Science & Humanities.
Finland Academic Research Science Publishers
ISSN: 2945-4492 (online) | (SJIF) = 8.09 Impact factor
Volume-14| Issue-11| 2026 Published: |22-11-2026|

RQ2. How does authentic video contribute to the development of intercultural awareness?

RQ3. Which learner, text, and task factors moderate these outcomes?

METHODOLOGY

2.1 Design

This study is a structured narrative review with a thematic synthesis. It does not report new primary data. The review design was chosen because the research questions span language outcomes, intercultural outcomes, and instructional design, which are studied with heterogeneous methods (experimental, quasi-experimental, meta-analytic, qualitative, and conceptual) that cannot be pooled statistically. Reporting was informed by the transparency principles of the PRISMA 2020 statement (Page et al., 2021), although the review is not a formal systematic review.

2.2 Sources and search strategy

Sources were purposively selected from the established literature rather than retrieved through an exhaustive database search. Selection targeted foundational theory and influential empirical studies and meta-analyses published in journals such as *Language Learning & Technology*, *System*, *Studies in Second Language Acquisition*, *ELT Journal*, and *Language Teaching*, and in major academic books. Topics guiding selection were *authentic video*, *captioned video*, *listening comprehension*, *vocabulary*, *intercultural competence*, and *task design*. Because selection was purposive, the review may under-represent recent or non-English-language studies (see Limitations).

2.3 Inclusion and exclusion criteria

Criterio	Included	Excluded
Materia	Video created for native-speaker audiences, or ed clips of such video	Video made solely for guage teaching
Learner	Second or foreign language learners of English olescent and adult)	First-language or young-child ulations
Focus	Language outcomes, intercultural outcomes, or ructional design	Technical studies unrelated to ning
Type	Peer-reviewed articles, meta-analyses, oretical books	Non-scholarly opinion pieces

2.4 Analysis

Included works were read in full and coded for (a) the language skill or cultural dimension addressed, (b) the type of video and viewing conditions, (c) accompanying tasks, (d) learner characteristics, and (e) reported outcomes. Codes

International Journal of Education, Social Science & Humanities.
Finland Academic Research Science Publishers
ISSN: 2945-4492 (online) | (SJIF) = 8.09 Impact factor
Volume-14| Issue-11| 2026 Published: |22-11-2026|

were grouped into themes following the six-phase approach to thematic analysis (Braun & Clarke, 2006): familiarization, initial coding, theme search, theme review, theme definition, and reporting. Themes were then mapped onto the three research questions.

RESULTS

Thematic analysis produced four themes, summarized in Table 1 and elaborated below.

Table 1. *Themes derived from the reviewed literature*

Theme	Core finding	Representative sources	Research question
1. Multimodal input and language gains	Captions and visual context support comprehension and incidental vocabulary learning	Montero Perez et al. (2013); Peters & Webb (2018); Sydorenko (2010)	RQ1
2. Video as cultural text	Video exposes pragmatic and cultural norms; reflection turns exposure into awareness	Byram (1997); Kramsch (1993); Chapple & Curtis (2000)	RQ2
3. Moderating factors	Proficiency, text difficulty, caption type, and cognitive load shape outcomes	Winke et al. (2010); Sweller (1988); Mayer (2009)	RQ3
4. Task design	Active, structured tasks convert viewing into language use	Berk (2009); Field (2008); Swain (1985)	RQ1–RQ3

3.1 Theme 1: Multimodal input and language gains (RQ1)

Research on captioned video provides the strongest evidence of language benefits. A meta-analysis of captioned video studies concluded that captions can support both listening comprehension and vocabulary learning (Montero Perez et al., 2013). Experimental work on listening points the same way: learners who watched videos with captions tended to comprehend more than those who watched without them (Winke et al., 2010), and English subtitles benefited EFL learners' listening comprehension in a film-based study (Hayati & Mohmedi, 2011). Earlier classroom work had already suggested that same-language subtitles help learners connect sound to written form (Vanderplank, 1988, 2016).

Vocabulary learning also occurs through viewing itself. Learners can pick up words incidentally from television programs, with learning related to how often a word occurs and how much support the image provides (Peters & Webb, 2018), and the combination of audio and written input has been linked to better word-form learning (Sydorenko, 2010). Webb (2015) argued that extensive viewing is a practical way to increase exposure beyond class time. Evidence on **speaking** is

International Journal of Education, Social Science & Humanities.
Finland Academic Research Science Publishers
ISSN: 2945-4492 (online) | (SJIF) = 8.09 Impact factor
Volume-14| Issue-11| 2026 Published: |22-11-2026|

thinner: studies generally treat video as input for oral work, such as discussion, retelling, and role-play, rather than testing speaking gains directly (Berk, 2009; Mekheimer, 2011).

3.2 Theme 2: Video as cultural text (RQ2)

Authentic video presents language embedded in social settings: forms of address, humor, politeness, gesture, and everyday behavior. This makes it a natural source for developing intercultural awareness as described by Byram (1997) and Kramsch (1993). Studies of film-based instruction report that learners respond positively to video and become more aware of cultural practices (Chapple & Curtis, 2000; Herron & Hanley, 1992). However, the literature also warns that exposure alone may reinforce stereotypes, because popular media portrays cultures selectively. Intercultural development therefore depends on guided observation, comparison with the learner's own culture, and critical discussion (Byram, 1997).

3.3 Theme 3: Moderating factors (RQ3)

Three groups of moderators recur. First, **learner proficiency**: lower-proficiency learners can be overwhelmed by speech rate and unfamiliar vocabulary, so the benefit of captions and shorter segments is greater for them, whereas advanced learners may profit more from reduced support (Winke et al., 2010; Montero Perez et al., 2013). Second, **text characteristics**: lexical coverage, speech rate, accent, and genre influence how much learners can understand (Peters & Webb, 2018; Rost, 2011). Third, **cognitive load**: simultaneous processing of audio, image, and text can either assist or overload working memory depending on the design (Mayer, 2009; Sweller, 1988).

3.4 Theme 4: Task design (RQ1–RQ3)

Across strands, authors emphasize that benefits rest on what learners do with the video. Practical syntheses recommend short clips, clear viewing purposes, repeated viewing, and structured pre-, while-, and post-viewing activities (Berk, 2009; Canning-Wilson, 2000; Field, 2008). Interactionist theory adds that tasks should push learners to negotiate meaning and produce output (Long, 1996; Swain, 1985). Designing authenticity into the *task*, not only the material, is a further recommendation (Mishan, 2005).

DISCUSSION

4.1 Interpretation of findings

The reviewed literature indicates that authentic video is a valuable but conditional resource. For RQ1, evidence is strongest for listening comprehension and vocabulary, where multimodal support, particularly same-language captions, helps learners process fast, unscripted speech (Montero Perez et al., 2013; Winke et

International Journal of Education, Social Science & Humanities.
Finland Academic Research Science Publishers
ISSN: 2945-4492 (online) | (SJIF) = 8.09 Impact factor
Volume-14| Issue-11| 2026 Published: |22-11-2026|

al., 2010). This is consistent with dual coding and multimedia learning accounts (Mayer, 2009; Paivio, 1986). For RQ2, video offers cultural content that textbooks rarely provide, yet intercultural awareness seems to depend on reflection rather than exposure (Byram, 1997). For RQ3, learner proficiency, text difficulty, and cognitive load determine whether support aids or hinders learning, and task design links all strands.

A notable pattern is that language-focused and culture-focused research rarely inform each other. Yet the same clip can serve both purposes: a scene of a service encounter, for instance, can be used to study discourse markers, listening strategies, and politeness norms. Integrating the two goals is both efficient and pedagogically coherent.

4.2 A three-phase framework for integration

Building on the themes above, the following framework organizes language and intercultural objectives around the familiar pre-, while-, and post-viewing sequence.

Table 2. *Integrated framework for using authentic video in ESL lessons*

Phase	Language-focused tasks	Intercultural-focused tasks
Pre-viewing	Preteach 5–8 key words; predict content from title or stills	Elicit learners' assumptions about the situation or culture portrayed
While-viewing	Gist viewing, then detailed viewing with captions (same-language) for lower levels; note-taking; selective replay	Observe and note behaviors, gestures, politeness, and values
Post-viewing	Retelling, discussion, role-play, written summary using target language	Compare with learners' own cultures; discuss stereotypes and variation; reflect on what could be misunderstood

The framework follows from the evidence: captions and pre-teaching manage cognitive load; repeated viewing deepens comprehension; and structured comparison and reflection convert cultural exposure into awareness (Byram, 1997; Canning-Wilson, 2000; Field, 2008).

4.3 Practical implications

Select carefully. Choose short clips (typically 2–5 minutes) matched to learners' proficiency and to a clear learning objective (Berk, 2009). The CEFR descriptors can help calibrate task difficulty (Council of Europe, 2020).

Scaffold, then fade. Begin with same-language captions and gradually reduce them as learners gain proficiency.

International Journal of Education, Social Science & Humanities.
Finland Academic Research Science Publishers
ISSN: 2945-4492 (online) | (SJIF) = 8.09 Impact factor
Volume-14| Issue-11| 2026 Published: |22-11-2026|

Require output. Follow every viewing with speaking or writing tasks so that input becomes acquisitional (Long, 1996; Swain, 1985).

Teach critical viewing. Present several portrayals of a culture and discuss how media constructs representations, to avoid reinforcing stereotypes.

Assess both goals. Use rubrics covering language accuracy and fluency alongside observation, comparison, and reflection skills.

4.4 Limitations

This review has several limitations. Sources were purposively selected rather than retrieved by an exhaustive systematic search, so relevant studies may be missing and selection bias is possible. The included studies vary in design, learner population, and outcome measures, which prevents quantitative synthesis. Much of the research is on short-term effects among university-level learners, and evidence on long-term retention, adolescent learners, and speaking outcomes is limited. Finally, the proposed framework is derived from theory and prior findings and has not been tested empirically.

4.5 Directions for future research

Future studies should test the framework in classroom experiments with pre- and post-tests of listening, vocabulary, speaking, and intercultural competence; compare caption conditions across proficiency levels; examine adolescent and low-proficiency learners; and use mixed methods (e.g., learner journals, interviews) to capture changes in attitudes toward other cultures.

5. Conclusion

Authentic video can strengthen both language proficiency and intercultural awareness in ESL classrooms, but only when it is matched to learners' level, supported by captions and pre-teaching, and embedded in tasks that require learners to speak, write, compare, and reflect. By integrating language and cultural objectives within a pre-, while-, and post-viewing sequence, teachers can use the same material to serve both aims. Empirical testing of this approach, particularly with adolescent learners and speaking outcomes, is the most pressing next step for researchers.

REFERENCES:

1. Berk, R. A. (2009). Multimedia teaching with video clips: TV, movies, YouTube, and mtvU in the college classroom. *International Journal of Technology in Teaching and Learning, 5*(1), 1–21.

2. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

3. Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.

4. Canning-Wilson, C. (2000). Practical aspects of using video in the foreign language classroom. *The Internet TESL Journal, 6*(11).

5. Chapple, L., & Curtis, A. (2000). Content-based instruction in Hong Kong: Student responses to film. *System, 28*(3), 419–433.

6. Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing.

7. Field, J. (2008). *Listening in the language classroom*. Cambridge University Press.

8. Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40*(2), 97–118.

9. Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology, 42*(1), 181–192.

10. Herron, C., & Hanley, J. (1992). Using video to introduce children to a foreign culture. *Foreign Language Annals, 25*(5), 419–426.

11. Kramsch, C. (1993). *Context and culture in language teaching*. Oxford University Press.

12. Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

13. 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.

14. Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

15. Mekheimer, M. A. (2011). The impact of using videos on whole language learning in EFL context. *Arab World English Journal, 2*(2), 5–39.

16. Mishan, F. (2005). *Designing authenticity into language learning materials*. Intellect Books.

17. 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.

18. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, Article n71. https://doi.org/10.1136/bmj.n71

19. Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

20. Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition, 40*(3), 551–577.