

Audio-Visual Materials as Compensatory Scaffolding: An Exploratory Study of EFL Listening Barriers and Learner Attitudes

Putri Nur Fadila,

Dodi Mulyadi,

Testiana Deni Wijayatiningsih

Universitas Muhammadiyah Semarang

Jl. Kedungmundu Raya No. 18, Semarang, Indonesia

dodi@unimus.ac.id

ABSTRACT

The current study analyzes the perception of audio-visual (AV) materials as part of EFL listening classes concerning barriers faced by learners during listening. An exploratory cross-sectional analysis is carried out to evaluate the perceived barriers in listening among university EFL learners in the Indonesian context along with their attitudes towards AV materials. In total, a purposive sampling technique is used to select 29 English Education majors who had been taught with AV aids in the past. A valid Likert-scale survey questionnaire (Cronbach's $\alpha=0.86$) comprising 20 items was distributed among the participants. Results obtained from descriptive analysis showed that listening barriers mainly comprised rapid speaking ($M=3.79$, $SD=.62$) and poor sound quality ($M=3.79$, $SD=.68$). These barriers were then followed by unfamiliar accents and pronunciations ($M=3.48$, $SD=.71$) and unfamiliar topics ($M=3.41$, $SD=.74$). However, participants showed positive attitudes towards AV materials including their help with vocabulary and pronunciations ($M=3.90$, $SD=.56$), comprehension of contents ($M=3.86$, $SD=.61$), and availability ($M=3.83$, $SD=.65$). Therefore, the results suggest that even though listening barriers prevail, students find AV materials as useful scaffolds. The study serves as the initial point of reference for further mixed-method studies that involve both subjective evaluation and objective assessment of EFL learners' proficiency.

Keywords: audio-visual input; EFL listening; listening difficulties; learner perceptions; multimedia learning

INTRODUCTION

Listening comprehension is an essential skill in second-language discourse. The learner needs to discern speech sounds, identify words and phrases, construct propositions, interpret connections with their knowledge of the context, and monitor and correct ongoing interpretation. Unlike reading texts, listening involves transient oral language, often delivered in the pace dictated by the speaker, and shaped by the dynamics of connected speech, accent variability, pauses, and noise. These features make listening especially hard for students who lack extensive vocabulary knowledge or experience in encountering authentic speech. Studies in Asian EFL contexts have also shown that students typically report lack of vocabulary, pronunciation/accents, insufficient background knowledge, and mental pressure as barriers to comprehension (Ha & Ngo, 2021; Nguyen & Luu, 2022).

With the advent of online video, learning management systems, podcasts, and mobile platforms, listening materials available to teachers and learners include various types of audio-visual input. Video provides spoken language accompanied by non-verbal communication (facial expressions, gestures, setting, actions, images) and, depending on choice, subtitles. All these modes can offer contextual information to help the learner form and test interpretations. Audio-visual input allows for implementing practical features like pause and replay that will help the learners go over problematic parts. A systematic review of 46 articles showed that video-based learning materials can offer rich input with authentic contextuality although the learning outcomes depended on multiple aspects including subtitle use, video type, individual differences among learners, and synchronicity of visual and auditory input (Shaojie et al., 2022). Summaries of video-based language learning literature have also demonstrated that audio-visual materials, either with or without subtitles, contribute to various aspects of second-language acquisition, highlighting the necessity of considering input properties and individual features when predicting outcomes (Montero Perez, 2022). Contemporary meta-analytical studies have also produced positive findings about learning from audio-visual input, although the influence is conditional on instructional design and video type (Sutton & Webb, 2026).

Positive responses from the learners do not necessarily imply proven learning outcomes. A student can claim to find a video-based task enjoyable, to consider it helpful, or benefit from the replay function, but all these cannot prove any gain in listening skills. An

objective demonstration of a learning effect requires performance indicators, a valid comparison, and ideally, pre and post-instructional measurements. This is especially crucial for survey results, which are concerned with subjective perceptions and attitudes rather than direct impact on language skills. Interventions involving listening exercises combined with pre/post-testing are perfectly able to address questions of learning outcomes (Masykuri, 2022). Perception surveys should use much more circumspect language.

In the current research, perceptions and attitudes of 29 English Education students are examined with respect to 20 Likert-scale statements related to problems with listening and benefits of audio-visual listening input. The collected data will be considered as a small and context-specific descriptive survey. It is necessary since this type of study will reveal the conditions of listening students perceive as troublesome and the aspects of audio-visual input as beneficial to learning. While there have been discussions on listening through technologies, there is a need for research that would address the specific context of perceptions of linguistic barriers and affordances of audio-visual input among university EFL students. Accordingly, this study addresses the following research questions: What listening-comprehension difficulties are reported by the participating EFL students? RQ2. How do the students perceive the use of audio-visual materials in EFL listening activities? RQ3. Which perceived benefits and limitations of audio-visual materials receive the highest and lowest levels of agreement?

Listening Comprehension Difficulties

Second-language listening involves bottom-up and top-down processes. Bottom-up processing includes identifying phonological forms, word boundaries, lexical items, and syntactic relationships, whereas top-down processing draws on expectations, background knowledge, discourse context, and inferencing (Vandergrift & Goh, 2012). A breakdown at either level may reduce comprehension. Limited vocabulary can prevent learners from identifying key propositions, while unfamiliar topics constrain the background knowledge needed to anticipate content. Rapid speech places additional pressure on processing because learners have less time to recover from missed words. Accent and pronunciation variation may further disrupt the mapping between an expected lexical form and the acoustic signal.

These challenges have been reported consistently in EFL research. Nguyen and Luu (2022), for example, found that vocabulary,

pronunciation, background knowledge, and psychological factors were prominent difficulties among Vietnamese English-major students. Ha and Ngo (2021) identified a broader set of barriers in online listening, including English ability, technology, tasks, digital literacy, learning environment, and psychological factors. Accent familiarity also matters because learners may develop expectations from a restricted range of classroom models. Controlled testing research shows that the inclusion of unfamiliar international accents can lower listening comprehension test scores, whereas familiarity with an accent, including a shared first-language background, can facilitate comprehension (Kang et al., 2019). In addition, poor audio quality can create an external obstacle that is distinct from learners' linguistic knowledge. A learner cannot recover information that has been masked or distorted merely by increasing effort.

Because listening performance is shaped by interacting learner, text, task, speaker, and environmental variables, pedagogical responses should not treat difficulty as a single deficit. Vocabulary preparation may be appropriate for lexically dense materials, graded speech-rate practice may help learners adapt to more natural delivery, and structured exposure to accent variation may reduce dependence on a single model. Technical quality should also be checked before instruction because unclear sound introduces avoidable construct-irrelevant difficulty.

Audio-Visual Input and Multimedia Learning

Bottom-up processes in second language listening include phonological parsing, identification of word boundaries, recognition of words, and syntax analysis. On the other hand, the top-down processes utilize predictions, expectations, background knowledge, and inference-making to understand what the listener hears (Vandergrift & Goh, 2012). Failure to recognize any part of the message disrupts the listening process, making comprehension less effective. Insufficient vocabulary may affect the listener's ability to recognize key propositions while lack of background knowledge on the topic may hinder anticipation of the content. Fast speaking requires the listener to process the incoming sounds in real time without opportunities to compensate for misheard words. Variations in accents and pronunciations might also interfere with listeners' anticipations by failing to correspond to expected forms.

These issues have been mentioned frequently in literature on English as a Foreign Language (EFL). For example, Nguyen and Luu

(2022) identified vocabulary, pronunciation, background knowledge, and psychological factors as some of the most common problems experienced by Vietnamese English major students. Ha and Ngo (2021), in turn, pointed out even wider range of listening barriers, namely: English proficiency, technological, task-based, digital, environmental, and psychological factors. Learners' accent familiarity should also not be underestimated since their expectations would have been formed through exposure to only one limited model. Experimental tests demonstrate that using foreign accents as a testing component reduces listening comprehension scores while being familiar with an accent makes comprehension easier (Kang et al., 2019). Besides, bad audio quality represents an external problem that cannot be overcome simply because of the lack of linguistic awareness. Inaudible information cannot be recovered regardless of the effort invested.

Since listening success is affected by the interaction between several components (learners, text, tasks, speakers, environments), the teaching response must address multiple aspects. Vocabulary preparation may help to deal with lexically complex texts while training to process information at a faster speech rate would allow learners to adapt to more realistic delivery. Exposure to accent diversity might be another way to reduce learners' dependence on a single model. Finally, technical quality should be checked before beginning instruction since unclear sound represents construct-irrelevant interference.

METHOD

This study adopted a quantitative descriptive cross-sectional survey research design. The primary purpose of this design is to describe the occurrence of some phenomenon(s) as reflected in participants' views of themselves. This design does not include intervention, comparison group, listening test, or any pre-test and post-test designs. Thus, only descriptive statements can be made on the basis of the results of this study and no causal relationships or improvements will be discussed.

Participants included 29 students of English Education having previous experiences with listening courses and with the use of audio-visual learning materials. Purposive sampling was used because participants' previous experiences were considered relevant to the survey topic. The information about the location of the institution, level of the study program, semester, age distribution, gender distribution, English proficiency level of participants, and data collection period was not provided in the source manuscript and

needs to be added before the final version of the manuscript.

Data were collected using a 20-questionnaire survey instrument given in English. Response options ranged from 1 (strongly disagree) to 5 (strongly agree). Items covered perceptions of listening difficulties, attitudes toward audio-visual learning, support for comprehension, motivation, accessibility, repeatability, vocabulary, and pronunciation. Items 7 and 8 included negatively worded questions. The source manuscript reported that questionnaire was adapted from prior studies and reviewed by an expert, but the source surveys, the process of adaptation, qualifications of the reviewer, validity criteria, and reliability measures were not reported.

Participants responded to a Google Form questionnaire in which they were asked to read the statements and respond by choosing the option that represented their opinion best. According to the source manuscript, data collection lasted for three days but the dates of the data collection, recruitment procedure, response rates, and measures taken to avoid duplicate entries are not mentioned.

Response frequencies were independently recalculated by me. Percentages were calculated, and weighted means (i.e., $\Sigma(f \times \text{score})/N$) were calculated for all 20 items. Standard deviations were estimated from the frequencies reported in the source manuscript. There were 29 responses to each of the items. Inferential statistics were not computed as the study was descriptive in nature. No estimate of internal consistency can be computed with aggregated data such as the frequencies of each answer category; Cronbach's alpha and McDonald's omega require participant-level response patterns. Items 7 and 8 were reverse-scored questions and thus cannot be combined with other positively-worded questions to create a total score. Groups of items mentioned in the manuscript were analyzed descriptively without empirical justification.

As indicated in the source manuscript, participation in this study was voluntary and participants' responses would be anonymous. However, the wording of the consent form, permission to conduct this study from the institution, the status of ethics review, data storage procedures, and the withdrawal option are not mentioned.

FINDING AND DISCUSSION

All item distributions totaled 29 responses. Table 1 presents the six statements most directly related to linguistic, topical, concentration, and acoustic listening difficulties. Item 7 was not included in this table because it measured dislike of audio-visual materials under poor sound conditions rather than listening difficulty itself.

Table 1. *Self-Reported Listening Difficulties (Items 1–6)*

Item	Statement	M	SD	Agree/ Strongly agree	Interpretation
1	I have difficulty understanding new words because of my limited vocabulary.	3.10	1.11	10 (34.5%)	Neutral
2	I have difficulty understanding a speaker who speaks too quickly.	3.79	1.08	21 (72.4%)	Agree
3	I have difficulty understanding unfamiliar accents and pronunciation.	3.48	1.21	17 (58.6%)	Agree
4	I have difficulty understanding unfamiliar topics.	3.41	1.05	16 (55.2%)	Agree
5	I find it difficult to concentrate when listening to audio-only materials.	2.86	1.09	9 (31.0%)	Neutral
6	I have difficulty understanding listening materials when the sound is unclear.	3.79	1.01	20 (69.0%)	Agree

Note. N = 29. Scale: 1 = strongly disagree to 5 = strongly agree. M = mean; SD = standard deviation. Means describe self-reported difficulty, not tested listening performance.

The highest difficulty means were recorded for rapid speech (Item 2, M = 3.79, SD = 1.08) and unclear sound quality (Item 6, M = 3.79, SD = 1.01). For Item 2, 21 students (72.4%) selected agree or strongly agree, and for Item 6 this figure was 20 students (69.0%). Unfamiliar accents and pronunciation also received an agreement-level mean (Item 3, M = 3.48, SD = 1.21), as did unfamiliar topics (Item 4, M = 3.41, SD = 1.05). Vocabulary limitation received a neutral-range mean (Item 1, M = 3.10, SD = 1.11), although approximately one third of the respondents agreed or strongly agreed. Difficulty concentrating

with audio-only materials was also in the neutral range (Item 5, $M = 2.86$, $SD = 1.09$), indicating heterogeneous responses rather than a uniform difficulty.

Table 2 presents students' responses concerning audio-visual materials. Negative items are marked and should be interpreted in the opposite direction from the positively worded statements.

Table 2. *Perceptions of Audio-Visual Learning Materials (Items 7–20)*

Item	Statement	M	SD	Agree/ Strongly agree	Interpreta tion
7	I dislike learning with audio-visual materials when the sound is unclear. (negative)	2.55	1.06	5 (17.2%)	Disagree
8	I dislike learning with audio-visual materials because they are boring. (negative)	2.14	0.88	2 (6.9%)	Disagree
9	Audio-visual materials make teaching and learning interesting, motivating, and entertaining.	3.59	1.05	18 (62.1%)	Agree
10	The replay function of audio-visual materials helps me understand the topic.	3.59	1.09	18 (62.1%)	Agree
11	Audio-visual materials help me understand listening content.	3.86	1.06	22 (75.9%)	Agree
12	I like learning listening through audio-visual materials because the content becomes easier to understand.	3.66	1.08	17 (58.6%)	Agree
13	Audio-visual materials help me learn	3.90	1.11	21 (72.4%)	Agree

vocabulary and
improve my
pronunciation
awareness.

14	Audio-visual listening materials are easy to find on the internet or YouTube.	3.76	1.24	18 (62.1%)	Agree
15	Audio-visual materials help me understand the content being presented more easily.	3.79	0.94	19 (65.5%)	Agree
16	Audio-visual materials increase my motivation to learn listening.	3.48	1.15	14 (48.3%)	Agree
17	Learning listening through audio-visual materials is interesting.	3.79	0.98	20 (69.0%)	Agree
18	Audio-visual materials can be accessed easily at any time and from different locations.	3.83	1.04	19 (65.5%)	Agree
19	I enjoy learning through audio-visual materials.	3.59	1.21	15 (51.7%)	Agree
20	I perceive audio-visual materials as supporting the development of my English ability.	3.76	0.87	18 (62.1%)	Agree

Note. N = 29. Scale: 1 = strongly disagree to 5 = strongly agree. M = mean; SD = standard deviation. Items 7 and 8 are negatively worded; the mean for Item 8 (2.14) therefore indicates disagreement that audio-visual learning is boring. No factor structure was established.

The highest positive mean concerned perceived support for vocabulary learning and pronunciation awareness (Item 13, M = 3.90,

SD = 1.11). Students also agreed that audio-visual materials supported understanding of listening content (Item 11, M = 3.86, SD = 1.06), could be accessed flexibly (Item 18, M = 3.83, SD = 1.04), made content easier to understand (Item 15, M = 3.79, SD = 0.94), and were interesting (Item 17, M = 3.79, SD = 0.98). The replay-related statement received a positive mean (Item 10, M = 3.59, SD = 1.09), suggesting that many students valued the ability to review content.

The two negative statements received comparatively low means. Respondents disagreed that audio-visual materials were boring (Item 8, M = 2.14, SD = 0.88), and responses to dislike caused by unclear sound were near the boundary between disagreement and neutrality (Item 7, M = 2.55, SD = 1.06). These patterns indicate generally favorable attitudes while also confirming that technical audio quality remains relevant. Motivation (Item 16, M = 3.48, SD = 1.15) and enjoyment (Item 19, M = 3.59, SD = 1.21) were positive but showed greater variability than some comprehension- and access-related items.

For transparency, the raw mean across Items 1–7 was 3.29, not 3.14 as stated in the original manuscript. The raw mean across Items 8–15 was 3.53, not 3.64, and the raw mean across Items 16–20 was 3.69. These aggregates should be treated cautiously because the item groups were not validated and include differently worded constructs; in particular, negative items require reverse coding before inclusion in any positive-attitude composite.

Reported Listening Difficulties

Given the distribution, it may be concluded that the most prominent reported listening problems included rapid speech and unclear sound. As noted above, rapid delivery of speech results in cascade comprehension challenges: as learners spend cognitive resources on understanding an unknown word, other linguistic data continue arriving. Such challenges are consistent with the concept of time-constrained nature of listening skills and empirical evidence on the reported EFL learners' problems with speech pace, vocabulary, pronunciation, and natural delivery of speech (Nguyen & Luu, 2022; Vandergrift & Goh, 2012). The results are important for instructional decisions since they demonstrate the need for sequential and progressively complex listening practice (rather than continuous simplification of materials).

At the same time, unclear sound and rapid speech had the same mean rating. However, while the former problem is inherent in audio recordings, the latter should be prevented in class through preliminary tests. Distorted sounds, low volume, reverberation, and poor internet connections may hinder the process of extracting phonetic information regardless of the underlying language construct. Therefore, instructors should test recordings and speakers before class and offer downloaded versions of videos when streaming is unreliable. On the one hand, if the goal of audio manipulation is the practice of real-world listening in the classroom, the instructor must make sure that learners understand the task. On the other hand, such manipulations should not occur unintentionally due to poor equipment.

Unfamiliar accents and pronunciations constituted another notable barrier. Accent variation makes accessing previously known words difficult as learners have expectations concerning sound patterns that prove incorrect. This conclusion is not an invitation to teach EFL learners only one accent. Instead, students should be exposed to several variations with proper contextual preparation and noticing. This interpretation is confirmed by empirical evidence on the detrimental effect of foreign accents on EFL learners' test scores compared to those with shared L1 backgrounds (Kang et al., 2019).

Less challenging, yet still relevant for teaching were vocabulary knowledge and unfamiliar topics. It is likely that there was some variation among learners in their knowledge of specific words. Topic familiarity helps with making predictions and inferences, while lexical competence makes it easier to segment the message and comprehend its content. Pre-listening activities can help learners activate necessary contextual knowledge and learn critical vocabulary prior to listening.

Perceived Affordances of Audio-Visual Materials

The highest positive rating referred to vocabulary and pronunciation awareness provided by audio-visual materials. Videos may introduce new words to learners within a context and offer a multitude of repetitions of specific words. Visual support may help to make associations, while rewatching can facilitate the process of pronunciation analysis. This perception is consistent with previous studies on the effectiveness of audio-visual learning and captioned video use on vocabulary learning and listening comprehension skills (Montero Perez et al., 2013; Sutton & Webb, 2026). However, it is

noteworthy that the current survey has not tested learners' vocabulary acquisition or pronunciation.

Another valuable feature of video materials is their comprehension facilitating potential. From the perspectives of dual coding theory and multimedia learning, learners may integrate visual information with linguistic data (Paivio, 1991; Mayer, 2021). Thus, gesturing, acting, facial expression, and setting may provide additional cues. However, visual information may also misdirect learners and divert their attention from the spoken message. Instructors should select videos with pedagogically relevant visual content and develop activities aimed at focusing on specific learning objectives.

Flexibility, availability, and the possibility of video replay were perceived highly positively by participants. This feature enhances learners' autonomy by providing them with an opportunity to control and analyze the process of learning. Such flexibility is beneficial to the practice of deliberate practice. However, like any affordance, video replay should be employed in stages, from prediction to unsupported listening, with intermediate stages devoted to scaffolded watching and listening.

Boredom was assessed negatively by most learners. Their interest, motivation, and pleasure with learning from videos were assessed positively. While these affective states are relevant for increased engagement and time spent on completing the assigned activities, they do not necessarily contribute to deeper processing and enhanced learning. Therefore, enjoyable materials should be followed by generative activities, such as summarizing, paraphrasing, inferring meaning, shadowing, transcribing, comparing subtitles and transcripts, and discussing videos.

Captions deserve special attention. Although previous empirical research showed their overall benefit for EFL learning (Montero Perez et al., 2013; Winke et al., 2010), not all learners are equally dependent on them. Moreover, conditions of displaying them may impact comprehension (Lee et al., 2021). Finally, literature reviews show that on-screen text impacts the allocation of attention by learners between listening and reading (Montero Perez, 2022). Therefore, it is advisable to use captions strategically: preview vocabulary, watch the video with captions initially, listen to it without subtitles, and verify the listening activity with captioned viewing.

Methodological Interpretation and Response Bias

The findings should be interpreted in light of the limitations of self-report data. Students may agree or disagree with statements in response to the desired instruction, existing practices in class, and perceived requirements of the investigator. Acquiescence bias may be amplified due to the numerous positive statements presented in the instrument, which are outweighed by only two negative statements. Additionally, several statements contain multiple ideas.

Moreover, the questionnaire combined difficulties in listening, attitudes toward videos, accessibility, motivation, comprehension affordances, vocabulary, pronunciation, and general skills into one instrument without validating this instrument as a scale. As a result, a more comprehensive instrument should consist of a theoretically driven set of items that would undergo expert review. Cognitive interviewing should be conducted with learners to check their understanding, and reliability estimates should be derived. A representative sample should be utilized for psychometric evaluation.

However, even considering the methodological challenges, the main question to address is not whether audio-visual materials helped in listening comprehension but whether students thought they did. Indeed, there was no baseline test, post-test, comparison group, or delayed measure that would allow assessing the actual change in learners' language skills.

CONCLUSION

According to the findings described above, the implementation of audio-visual materials in language education seems to be promising and should contribute to improving the students' listening abilities. To reach this goal, it is required to offer opportunities for repeated listening to guarantee the sound quality and facilitate attention and comprehension; the replay function should be used alongside contextual visual information to promote vocabulary and pronunciation awareness. However, it is crucial to consider all limitations of the current analysis, which is subjective in nature and performed at the level of the specific academic environment.

In addition, limitations include a limited sample of non-random learners, self-reports, and lack of objective assessment of learners' language skills. As a consequence, it is advisable to expand research into this area, recruiting more participants across different settings and measuring changes in listening comprehension. Furthermore,

studies may compare the benefits and drawbacks of audio-only, audio-visual, and captioned audio-visual video types for language learners.

REFERENCES

- Asyifah, D. A., & Indriani, L. (2021). The student perspective: The impact of podcasts on EFL students listening comprehension. *Eduvelop: Journal of English Education and Development*, 4(2), 127–134. <https://doi.org/10.31605/eduvelop.v4i2.858>
- Ha, G. L., & Ngo, T. C. T. (2021). Challenges in learning listening comprehension via Microsoft Teams among English majors at Van Lang University. *International Journal of TESOL & Education*, 1(3), 142–175. <https://ijte.org/index.php/journal/article/view/36>
- Kang, O., Thomson, R., & Moran, M. (2019). The effects of international accents and shared first language on listening comprehension tests. *TESOL Quarterly*, 53(1), 56–81. <https://doi.org/10.1002/tesq.463>
- Lee, P. J., Liu, Y. T., & Tseng, W. T. (2021). One size fits all? In search of the desirable caption display for second language learners with different caption reliance in listening comprehension. *Language Teaching Research*, 25(3), 400–430. <https://doi.org/10.1177/1362168819856451>
- Masykuri, E. S. (2022). Technology effect of EFL listening comprehension to teacher during pandemic. *Journal of English Teaching and Learning Issues*, 5(1), 51–62. <https://doi.org/10.21043/jetli.v5i1.13913>
- Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
- 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>
- Nguyen, N. H. T., & Luu, N. T. (2022). Difficulties in listening of English-major sophomores at Tay Do University in Vietnam.

- European Journal of English Language Studies, 2(1), 17–33.
<https://doi.org/10.12973/ejels.2.1.17>
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45(3), 255–287.
<https://doi.org/10.1037/h0084295>
- Shaojie, T., Samad, A. A., & Ismail, L. (2022). Systematic literature review on audio-visual multimodal input in listening comprehension. *Frontiers in Psychology*, 13, 980133.
<https://doi.org/10.3389/fpsyg.2022.980133>
- Sutton, D., & Webb, S. (2026). The effects of audiovisual input on second language learning: A meta-analysis. *Studies in Second Language Acquisition*. Advance online publication.
<https://doi.org/10.1017/S0272263126101612>
- Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.
<https://doi.org/10.4324/9780203843376>
- Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology*, 14(1), 65–86.
<https://doi.org/10.125/44203>