International Journal of Learning, Teaching and Educational Research
Vol. 25, No. 7, pp. 265-301, July 2026
https://doi.org/10.26803/ijlter.25.7.13
Received Apr 10, 2026; Revised Jun 12, 2026; Accepted Jun 22, 2026

Video-Based Shadowing in a Content and Language Integrated Learning (CLIL) Framework: A Study of Indonesian Learners of Japanese as Second Language (JSL) Perceived Listening, Pronunciation, and Speaking Development

Dian Bayu Firmansyah* , **Octaria Putri Nurharyani**,
Monika Herliana, **Fasha Hegga Sholtanaju**
Universitas Jenderal Soedirman
Purwokerto, Indonesia

Lisda Nurjaleka
Waseda University
Tokyo, Japan

Abstract. Despite growing research on YouTube-assisted language learning and shadowing, little is known about how video-based shadowing operates within a Content and Language Integrated Learning (CLIL) in Japanese as a Second Language (JSL) contexts. This study examined the extent to which video-based shadowing enhances Indonesian JSL learners' perceived oral proficiency within a CLIL framework and explored the perceived benefits and challenges of its implementation across linguistic, affective, and cross-cultural dimensions. Employing a convergent mixed-methods design, 57 intermediate-level Indonesian JSL learners completed Likert-scale questionnaires, open-ended responses and semi-structured interviews while using authentic content from the Speak Japanese Naturally YouTube channel. Results indicated strong positive perceptions across listening (M=4.69–4.94), speaking (M=4.64–4.75), shadowing performance (M=3.92–4.86), and satisfaction (M=3.67–4.81) subscales. Furthermore, findings identified speaking improvement (n=17), listening comprehension (n=15), and pronunciation accuracy (n=12) as the most frequently cited benefits. Joint display integration further revealed that high learner satisfaction (M=4.81) coexisted with cognitive overload from excessive speech rate (n=18) and affective strain (n=10). The findings suggest that video-based shadowing within a CLIL framework

Citation:
Firmansyah, D. B., Nurharyani, O. P., Herliana, M., Sholtanaju, F. H., Nurjaleka, L. (2026). Video-Based Shadowing in a Content and Language Integrated Learning (CLIL) Framework: Study of Indonesian Japanese as Second Language (JSL) Learners' Perceived Listening, Pronunciation, and Speaking Development. *International Journal of Learning, Teaching and Educational Research*, *25*(7), 265–301. https://doi.org/10.26803/ijlter.25.7.13

*Corresponding author: Dian Bayu Firmansyah; *dbayuf@unsoed.ac.id*

© The Authors
This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

constitutes a theoretically coherent and replicable strategy for spoken Japanese development in resource-constrained Southeast Asian JSL contexts. The study also offers implications for curriculum sequencing, assessment design, teacher professional development, and the future integration of AI-assisted pronunciation feedback tools.

Keywords: video-based shadowing; spoken proficiency in Japanese; Content and Language Integrated Learning (CLIL); autonomous language learning; YouTube

1. Introduction

Pronunciation and listening comprehension are fundamental aspects of second language (L2) acquisition, contributing directly to learners' communicative competence and overall intelligibility (Levis, 2020; Whitworth & Rose, 2025). Learners who struggle with these skills may experience communication breakdowns, misunderstandings, and reduced willingness to communicate (Yates, 2011; Zielinski, 2011; Zoss, 2016). Despite their importance, pronunciation and listening instruction are often underrepresented in language curricula due to difficulties in aligning them with communicative teaching approaches (Derwing, 2010; Murphy, 2014).

Amid these challenges, the shadowing technique has emerged as a promising approach to developing learners' listening and pronunciation skills (Tamai, 1997; Hamada, 2016; Hamada, 2019). Originally designed for simultaneous interpreter training (Lambert, 1992), shadowing involves repeating speech almost simultaneously with auditory input, a process that activates attention in the working memory and engages both listening and speaking channels (Tamai, 1992; 1997; Kurata, 2007). This dual-task mechanism differentiates shadowing from traditional listen-and-repeat drills, fostering greater focus on phonological and prosodic features (Hamada & Suzuki, 2022). Empirical evidence supports its effectiveness in enhancing listening comprehension, pronunciation, and learner motivation across English as Foreign Language (EFL) and Japanese as Second Language (JSL) contexts (Hamada, 2011; Hamada, 2015; Kadota, 2012; Shiki et al., 2010; Mochizuki, 2006; Omar & Umehara, 2010).

However, the integration of shadowing into different L2 learning contexts remains challenging. In contexts outside Japan, where communicative language teaching is widely adopted, the repetitive and mechanical nature of shadowing may be perceived as incompatible with meaning-focused instruction and could even increase learner anxiety or fear of negative evaluation (Horwitz et al., 1986; Markee in Sumiyoshi & Svetanant, 2017; Hamada, 2015). These cultural and pedagogical differences underscore the need for adaptive applications of shadowing that balance its mechanical rigour with communicative goals. When carefully implemented—such as through individualized or homework-based practice—shadowing can mitigate performance anxiety while reinforcing listening comprehension and pronunciation accuracy (Sumiyoshi & Svetanant, 2017). This issue is particularly relevant for Indonesian JSL learners, who are

generally taught within communicative teaching traditions but still need sustained exposure to develop natural spoken Japanese.

In recent years, digital platforms—particularly YouTube—have emerged as influential tools for autonomous language learning, offering multimodal, authentic input accessible beyond the classroom (Van et al., 2021; Mutiarani, 2021; Satriani, 2023; Zeng, 2020). For JSL learners in Indonesia, where opportunities for immersive exposure to natural Japanese speech are limited, YouTube provides a practical and scalable medium for shadowing practice.

The integration of shadowing into the Content and Language Integrated Learning (CLIL) framework presents an innovative pedagogical direction. CLIL emphasizes learning content through a foreign language, thereby fostering linguistic proficiency and intercultural awareness simultaneously. In Japanese language education, CLIL can be realized through culturally embedded materials such as authentic YouTube videos, which allow learners to engage with both linguistic and cultural dimensions of communication (Coyle et al., 2010). This alignment between shadowing's phonological focus and CLIL's emphasis on meaningful content makes video-based shadowing via YouTube a particularly apt strategy for developing spoken proficiency and cross-cultural competence among JSL learners.

In addition to serving as a source of authentic input, YouTube also facilitates active engagement and communication. Its interactive features, such as comment sections, enable learners to exchange ideas, practice writing, and reflect on linguistic forms in meaningful contexts (Coyle et al., 2010; Zulkflee et al., 2022). Such participatory functions can complement speaking practice by reinforcing learners' understanding of language use in socially and culturally appropriate ways. In the Japanese learning context, where politeness levels, turn-taking, and pragmatic cues are vital, these interactions can deepen learners' awareness of communicative norms.

Furthermore, empirical findings suggest that using YouTube as a language learning medium contributes significantly to improving speaking proficiency. Students exposed to video-based materials show measurable progress in vocabulary acquisition, grammatical accuracy, pronunciation, and the ability to generate ideas related to the content of the videos. These improvements indicate that video-based learning through YouTube not only enhances linguistic performance but also promotes contextualized thinking and cultural understanding, both of which are central to effective spoken communication in Japanese.

Despite the growing body of research on shadowing (Hamada, 2019; Kadota, 2012) and YouTube-based language learning (Van et al., 2021; Mutiarani, 2021; Satriani, 2023), empirical studies specifically examining video-based shadowing within a CLIL framework for Indonesian JSL learners remain scarce. Most existing studies concentrate on EFL contexts or on native-speaker-oriented environments in Japan, leaving the intersection of shadowing, authentic video input, and cross-

cultural communicative competence in JSL settings underexplored. Furthermore, no prior study has operationalized shadowing within a CLIL design to simultaneously target phonological accuracy, listening comprehension, and pragmatic awareness among Indonesian university-level learners of Japanese.

The central problem of this study lies in the absence of evidence-based pedagogical strategies that integrate shadowing with meaning-focused, content-rich instruction for Indonesian JSL learners. Although communicative language teaching (CLT) is widely adopted in Indonesian higher education (Haryono, 2019), its emphasis on fluency over phonological accuracy tends to marginalize listening and pronunciation training (Derwing, 2010; Murphy, 2014). Meanwhile, available digital resources, particularly YouTube, remain largely unstructured in their classroom applications, without a systematic framework to maximize their potential for spoken proficiency development. This gap between available tools and principled pedagogical integration constitutes the core problem this study seeks to address.

This study has four objectives: (1) to measure the extent to which video-based shadowing practice within a CLIL framework improves Indonesian JSL learners' listening comprehension and speaking performance, as reflected in their self-reported perceptions; (2) to identify and analyze the positive and negative aspects of video-based shadowing as perceived by Indonesian JSL learners across linguistic, affective, and cultural dimensions; (3) to explore learners' perceptions of cross-cultural and pragmatic awareness as an emergent outcome of video-based shadowing within the CLIL framework; and (4) to generate evidence-based implications for educators and curriculum designers seeking to integrate technology-mediated shadowing into JSL teaching at the university level.

In light of these issues and the studies reviewed above, this study addresses the following research questions:

1. To what extent do Indonesian JSL learners perceive video-based shadowing within a CLIL framework as improving their listening comprehension, pronunciation accuracy, and speaking fluency?
2. What benefits and challenges do Indonesian JSL learners perceive in video-based shadowing, particularly in relation to linguistic gains, affective responses, cognitive demands, and cross-cultural or pragmatic awareness, and how do these factors affect the perceived sustainability of independent shadowing practice?

2. Literature Review

Three interrelated theoretical constructs underpin this study: Content and Language Integrated Learning (CLIL), shadowing, and learner autonomy. CLIL is defined as a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language, with the goal of promoting language proficiency alongside subject knowledge (Coyle et al., 2010; Dalton-Puffer, 2011). Shadowing, in turn, is a technique in which learners simultaneously or near-simultaneously repeat spoken input, originally developed for interpreter training and subsequently adapted for L2 pronunciation

and listening development (Hamada, 2019; Kadota, 2012). Learner autonomy refers to the capability of managing their own learning and that this could quite readily take place outside the classroom (Lewis, 2014). The convergence of these three constructs in the present study is not incidental: CLIL supplies the content-rich, culturally embedded framework; shadowing provides the phonological processing mechanism; and learner autonomy is enabled by the YouTube platform, which allows self-paced, repeated engagement with authentic input. Together, they constitute a theoretically coherent model for spoken proficiency development in JSL contexts.

2.1 YouTube as a Multimodal Medium for Autonomous Language Learning: Affordances and Limitations

The rapid expansion of digital technologies has fundamentally reshaped contemporary language learning environments, positioning online platforms as central mediating tools in both formal and informal education. Among these platforms, YouTube has emerged as one of the most influential digital resources due to its accessibility, multimodal affordances, and widespread adoption among learners. In the context of JSL learning in Indonesia, where direct access to authentic Japanese speech is limited outside the classroom, YouTube represents a particularly salient medium for bridging this exposure gap (Nasution, 2019; Nofrika, 2019). In this context, YouTube serves not only as a source of entertainment but also as a repository of educational content that supports language acquisition at all proficiency levels.

A substantial body of research highlights the pedagogical potential of YouTube in facilitating foreign language learning, particularly in enhancing learner engagement and comprehension. The platform provides access to authentic, multimodal input, allowing learners to simultaneously process auditory and visual information, which contributes to improved understanding of linguistic forms and contextual usage (Brinton, 2001; Almurashi, 2016). Empirical studies consistently report positive learner perceptions of YouTube, indicating that it fosters enthusiasm, promotes active participation, and supports independent learning (Aprianto, 2020; Putri et al., 2020).

Similarly, learners have been found to benefit from YouTube in completing academic tasks and improving overall language proficiency, particularly in English as a Foreign Language (EFL) context (Sakkir et al., 2020; Hia, 2021). These findings suggest that YouTube serves not merely as a supplementary tool but as a meaningful component of contemporary language learning ecosystems. However, it is important to note that these findings derive predominantly from EFL contexts; their transferability to JSL settings, where phonological and prosodic demands differ substantially from English, warrants careful consideration (Whitworth & Rose, 2025).

In terms of language skill development, existing studies demonstrate that YouTube significantly enhances receptive and productive skills. For instance, exposure to YouTube videos has been shown to improve listening comprehension by providing contextualized and multimodal input (Qomariyah et al., 2021), while repeated engagement with authentic speech content enhances learners'

pronunciation accuracy, fluency, and confidence in speaking (Mutiarani & Rusiana, 2021; Alkathiri, 2019). Furthermore, YouTube supports vocabulary, grammar, and pronunciation development through interactive, repetitive practice (Chhabra, 2012; Putri, 2022). While writing skill development via YouTube has also been documented (Sianna et al., 2018), the present study focuses specifically on oral modalities such as listening and speaking, where the convergence with shadowing is most theoretically defensible.

Beyond skill development, YouTube also plays a crucial role in fostering learner autonomy and self-regulated learning. The platform's vast and diverse content enables learners to select materials that align with their individual interests, proficiency levels, and learning goals, thereby facilitating personalized learning experiences (Hughey, 2020). This flexibility enhances learner motivation and encourages sustained engagement, as students are more likely to invest effort in learning activities that they perceive as relevant and enjoyable (Gray & Diloreto, 2016; Daeli, 2024). Moreover, the availability of authentic language input across various sociocultural contexts contributes to the development of pragmatic and intercultural competence (Yuyun & Simamora, 2021; Danial, 2022). This dimension is particularly significant in the present study, as the selected channel "Speak Japanese Naturally" embeds cultural and pragmatic content within its speech, offering learners exposure to socially situated Japanese beyond formal register.

The interactive features of YouTube further extend its pedagogical value by enabling active learner engagement and social interaction. Learners employ a range of strategies when engaging with video content, including notetaking, pausing and replaying segments, and utilizing subtitles to enhance comprehension and retention (Jin & Webb, 2021; Holmes & Tuomi, 2022; Fitri & Ma'rifah, 2022). Additionally, participation in comment sections and live chat discussions allows learners to engage in meaningful communication, transforming passive viewing into interactive learning experiences (Zhang & Lee, 2023; Byun et al., 2023). These interactions not only facilitate language practice but also create opportunities for collaborative learning and community building, which are essential components of effective language acquisition (Bárcena et al., 2015). In the context of shadowing practice, the ability to replay and isolate specific utterances is especially relevant, as it allows learners to focus on prosodic and phonological targets that would otherwise be inaccessible in real-time speech (Hamada, 2019).

Despite its numerous advantages, using YouTube for language learning is not without limitations. One of the primary challenges is the overwhelming volume of available content, which can lead to information overload and difficulty in identifying high-quality instructional materials (Shanmugasundaram & Tamilarasu, 2023; Auxier & Vitak, 2019). The absence of systematic content curation means that learners may encounter inaccurate or pedagogically unsound materials, potentially hindering their learning progress (Zengin & Onder, 2021; Shoufan & Mohamed, 2022). Furthermore, the platform's entertainment-oriented nature may distract learners from educational objectives, reducing the efficiency

of learning activities and undermining sustained engagement (Khan et al., 2024). The lack of structured guidance, feedback, and accountability in autonomous learning contexts also poses challenges for maintaining motivation and consistency in learning (Al Harrasi, 2023). These limitations are not trivial, and represent a key rationale for embedding YouTube within a structured pedagogical framework, specifically CLIL rather than relying on incidental or unguided viewing.

2.2 Shadowing in L2 Acquisition: Theoretical Foundations, Evidence and Pedagogical Tensions

Shadowing as an L2 learning technique has attracted sustained empirical attention, particularly in Japanese EFL and JSL contexts (Hamada, 2019; Kadota, 2012; Whitworth & Rose, 2025). At its core, shadowing activates the phonological loop of working memory, requiring simultaneous attention to incoming speech and its immediate reproduction (Whitworth & Rose, 2025). This dual-task processing distinguishes shadowing from passive listening or delayed repetition drills and is theorised to accelerate automatization of bottom-up phonological processing (Kadota, 2012). Empirical evidence supports its effectiveness in improving listening comprehension (Kato, 2009), pronunciation accuracy (Bovee & Stewart, 2008), and learner motivation (Hamada, 2011; Hamada, 2015).

However, the evidence base for shadowing is not without contradictions. Several studies report that learners, particularly beginners, may experience cognitive overload when simultaneously processing and reproducing unfamiliar speech, potentially reducing comprehension rather than enhancing it (Hamada, 2015; Shiki et al., 2010). Furthermore, concerns have been raised about the mechanical nature of shadowing and its compatibility with communicative language teaching (CLT) principles; critics argue that repetitive, form-focused practice may undermine meaning-making and communicative confidence if not carefully scaffolded (Markee, cited in Sumiyoshi & Svetanant, 2017). These tensions suggest that shadowing is not a universally effective technique but one whose effectiveness is contingent on proficiency level, task design, and instructional context.

Alternative and complementary approaches to shadowing exist also merit brief consideration. Oral recitation (Newton & Nation, 2021) and extensive listening (Renandya & Farrell, 2011) offer meaning-focused approaches to building fluency, though they lack shadowing's precision in targeting prosodic imitation. The present study does not claim that shadowing is superior to these alternatives; rather, it posits that when combined with authentic content delivery via YouTube within a CLIL framework, video-based shadowing provides a pedagogically distinctive and theoretically grounded approach that neither incidental listening nor form-focused drilling alone can replicate.

2.3 CLIL as a Framework for Content-Driven Phonological Development: Rationale for Integration with Video-Based Shadowing

A central theoretical question for this study concerns how CLIL and shadowing, two approaches that operate at different levels of language processing, can be coherently integrated. CLIL, by design, foregrounds meaning and content,

engaging learners at the level of discourse and communicative intent (Coyle et al., 2010; Dalton-Puffer, 2011). Shadowing, conversely, operates at the sub-lexical and phonological level, targeting prosody, rhythm, and segmental accuracy (Kadota, 2012; Hamada, 2019). The apparent tension between these orientations requires theoretical justification rather than assumption.

The integration of CLIL and shadowing is theoretically grounded in contemporary input-based accounts of second language acquisition. Lichtman and VanPatten (2021) argue that comprehensible input remains the principal data for the acquisition of language, affirming that communicatively embedded input is a cornerstone shared across all major SLA theories — an insight that extends and refines, rather than discards, earlier foundational thinking in the field. Long (2015) further develops this position by arguing that acquisition is optimised when input is made both comprehensible and purposeful through meaning-focused instructional tasks, a condition that CLIL satisfies by embedding target-language exposure within authentic, content-driven activity.

Within this framework, shadowing functions as a form-focused complement: by requiring simultaneous phonological reproduction of heard input, it activates the bottom-up decoding processes necessary to consolidate what learners have partially comprehended at the meaning level. This combination reflects what Schmidt (1990) termed noticing, the conscious registration of target forms argued to be a prerequisite for acquisition. By shadowing authentic, culturally embedded video content, learners are simultaneously exposed to content worth noticing and trained to notice its phonological form, creating a dual-channel reinforcement loop in which meaning-focused input and form-focused processing mutually support one another.

Furthermore, CLIL's emphasis on the "4Cs" (Content, Communication, Cognition, and Culture; Coyle et al., 2010) maps directly onto the dimensions targeted in this study: content is supplied by the YouTube channel's thematic videos on Japanese daily life; communication is practised through shadowing-based speaking; cognition is engaged through pre- and post-viewing comprehension tasks; and culture is addressed through the pragmatic and sociolinguistic features embedded in natural Japanese speech. This alignment provides a more principled basis for the study than a simple additive combination of tools and distinguishes the present framework from studies that use YouTube merely as a supplementary resource without systematic pedagogical design.

3. Methodology

3.1 Research Design

This study employed a convergent mixed-methods design (Katz-Buonincontro, 2024), in which quantitative data from Likert-scale questionnaires and qualitative data from open-ended responses were collected concurrently, analyzed independently, and subsequently merged through a joint display to compare, contrast, and integrate findings across both strands. The integration of both strands at the interpretation phase allowed for triangulation, strengthening the credibility and completeness of the findings.

3.2 Participants

A total of 57 intermediate-level learners of Japanese (male: 21, female: 36) participated in this study. Participants ranged in age from 19 to 20 years and had been studying Japanese for a minimum of two years at the time of data collection. All participants were enrolled in Japanese language programmes at the undergraduate level and had attained at least Japanese Language Proficiency Test (JLPT) N4 proficiency.

Participants were selected using purposive sampling (Campbell et al., 2020), targeting learners who were actively enrolled in Japanese language courses at the time of the study. It should be acknowledged, however, that this sampling approach introduces potential selection bias: participants drawn from formal university programmes may demonstrate higher motivation and digital literacy than the broader population of Indonesian JSL learners. Furthermore, the study does not include beginner or advanced learners, limiting the generalizability of findings to intermediate-level university contexts.

3.3 Data Collection and Analysis

To collect data, multiple instruments were employed. Questionnaires were used to gather general information about learners' experiences and attitudes, containing five items in each of four categories: (1) listening competence; (2) speaking competence; (3) shadowing practice behaviour and skill development; and (4) overall satisfaction with the shadowing activity. Items were measured on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with negatively phrased items reverse-scored for analysis. The instrument was adapted from Spiller et al. (1990) and Hamada (2015) (See Appendix 1). Open-questions (See Appendix 1) and semi-structured interviews (See Appendix 2) allowed for deeper exploration of individual perspectives, perceived challenges, and benefits. Reflective journals, while collected as part of the broader research project, were not included in the present analysis due to the study's focus on immediate post-session perceptions. Questionnaires were administered anonymously, and participants were explicitly informed that there were no correct or incorrect responses.

Quantitative analysis: Descriptive statistics were computed for all Likert-scale items, including means (M), standard deviations (SD), and frequency distributions of agreement levels. Data were processed using SPSS version 26. No inferential statistics were applied, as the study did not seek to test directional hypotheses; rather, the quantitative strand aimed to map the overall pattern of learner perceptions across the measured dimensions.

Qualitative analysis: Open-ended questionnaire responses and semi-structured interview transcripts were analyzed using Braun and Clarke's (2006) six-phase thematic analysis procedure. Themes were identified inductively across both sources to ensure that qualitatively reported experiences were not contingent on any single instrument.

A potential methodological concern in this study is the possibility of experimenter demand effects and the Hawthorne effect (Zizzo, 2010; Adair, 1984): participants

who are aware of being observed or who seek to respond in socially desirable ways may overreport positive perceptions of the shadowing intervention. To mitigate this, questionnaires were administered anonymously, participants were explicitly informed that there were no correct or incorrect responses, and the research team was not present during questionnaire completion. Nonetheless, self-report bias cannot be entirely eliminated and should be considered when interpreting the quantitative findings.

3.4 Research Instruments and Materials

The questionnaire was adapted from previously validated instruments developed by Spiller et al. (1990) and Hamada (2015). Minor contextual modifications were made to align the items with the Indonesian JSL context. To assess the reliability of the questionnaire, Cronbach's alpha was computed as a measure of internal consistency. The overall 20-item instrument demonstrated excellent reliability ($\alpha = .914$). Subscale reliabilities were good for Listening Competence ($\alpha = .849$) and acceptable for Speaking Competence ($\alpha = .762$). The Shadowing Practice Performance and Satisfaction subscales yielded lower coefficients ($\alpha = .712$ and $\alpha = .747$, respectively), attributable primarily to near-ceiling response distributions that restricted inter-item covariance, a known limitation of Cronbach's alpha under low-variance conditions (Sijtsma, 2009).

For the instructional materials, learners engaged with culturally rich Japanese YouTube videos from the Speak Japanese Naturally channel. Videos were selected based on four criteria: (a) speech rate appropriate for intermediate learners (b) clear audio quality without significant background noise, (c) thematic content embedding Japanese cultural and pragmatic norms, and (d) duration of five to ten minutes to allow repeated shadowing within a single session.

3.5 Video-Based Shadowing Practice Steps

To maximize the effectiveness of the shadowing practice, respondents in this study followed a series of procedures adapted from the shadowing training steps proposed by Shiki et al. (2010). The process consisted of the following stages:

1. Respondents selected one of ten videos from the Speak Japanese Naturally YouTube channel, each presenting a different topic prepared for this research. Participants were encouraged to choose a theme of personal interest to maintain engagement and reduce cognitive burden during practice.
2. Respondents who selected the same video theme were grouped to facilitate shared understanding of the video content and to allow for peer feedback on their shadowing performance.
3. In the first session, respondents performed shadowing (1st attempt) using the chosen video without prior knowledge of its content or meaning, allowing them to focus solely on listening and speech reproduction.
4. After the initial practice, respondents were given time to comprehend the video's linguistic and cultural content. This stage was intended to deepen their understanding of both meaning and contextual aspects embedded in the material.
5. Once the content had been understood, respondents conducted a second attempt at shadowing, integrating their newly gained comprehension into the practice.

6. Following the second attempt at shadowing practice, participants engaged in content-based reflection and comprehension checks to evaluate their understanding of the selected video.
7. The procedure concluded with a third attempt and final round of shadowing, designed to observe changes and progress from the earlier sessions. The shadowing phase ended with a feedback discussion focusing on respondents' performance and learning outcomes.

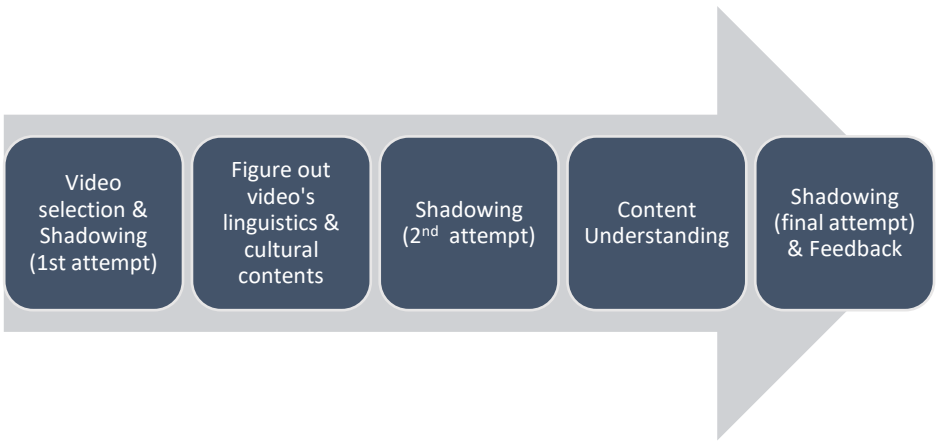

Figure 1: Video-based Shadowing Practice Steps

3.6 Ethical Considerations

Before the study commenced, formal approval was secured from the relevant institutional ethics board. Participants were informed about the purpose and procedures of the research, and they were assured that their responses would remain confidential and anonymous. They were also made aware that participation was entirely voluntary and that they could discontinue their involvement at any stage without penalty. Written consent was obtained from all participants before data collection, and the research was conducted in accordance with established ethical standards for studies involving human participants. All qualitative data were anonymized prior to analysis, and no identifying information was retained in the dataset.

4. Results and Findings

4.1 Participants' Responses to Questionnaire Items

Table 1 summarizes the frequency and percentage of responses for the listening competence questionnaire items of the survey. The responses, 'Strongly Agree', 'Agree', 'Disagree', and 'Strongly Disagree', were combined to simplify the table; however, the actual five-point scale scores were used, and negatively phrased items were reverse-scored to ensure consistent positive scoring for statistical analyses.

The results indicate that students overwhelmingly perceived video-based shadowing as highly effective for enhancing their listening comprehension in Japanese. Across all positively worded items (Items 1–4), more than 73% of respondents selected agree or strongly agree, with mean scores ranging from 4.69 to 4.94. Item 4, in particular, shows the strongest endorsement (M = 4.94),

suggesting that most students felt noticeable improvement in their listening ability after engaging in video-based shadowing. These consistently high ratings demonstrate a strong belief among learners that the technique promotes listening development, especially when practiced regularly.

Table 1: Listening Competence

No	Statement	Strongly Agree	Agree	Disagree	Strongly Disagree	Mean (SD)
1	Video-based shadowing is effective in improving listening skills.	44 (77.1%)	7 (12.2%)	5 (8.7%)	1 (1.7%)	4.69 (1.091)
2	Listening skills improve if video-based shadowing improves.	42 (73.6%)	12 (21.0%)	3 (5.2%)	0 (0%)	4.72 (0.914)
3	Listening skills improve with more practice in video-based shadowing.	43 (75.4%)	10 (17.5%)	3 (5.2%)	1 (1.7%)	4.72 (1.031)
4	Became better at listening after practicing video-based shadowing.	49 (85.9%)	5 (8.7%)	3 (5.2%)	0 (0%)	4.94 (0.860)
5	Don't think video-based shadowing is good for listening skills.	0 (0%)	1 (1.7%)	12 (21.0%)	44 (77.2%)	4.78 (1.098)

Table 2 summarizes the frequency and percentage of responses for the speaking competence questionnaire items of the survey. The data indicate strong learner agreement that video-based shadowing is effective for improving pronunciation and speaking skills. Items 6–8 received high levels of endorsement, with 77–81% of respondents agreeing or strongly agreeing that shadowing enhances pronunciation (Item 6) and contributes to overall speaking development (Items 7 and 8).

The mean scores for these items, ranging from 4.69 to 4.75, indicate consistently positive perceptions of shadowing as a technique that supports oral production through repeated and guided practice with authentic audiovisual input. Although Item 9 shows slightly lower agreement (M = 4.31), a clear majority (64.9%) reported noticeable improvement in their pronunciation after participating in video-based shadowing activities.

Table 2: Speaking Competence

No	Statement	Strongly Agree	Agree	Disagree	Strongly Disagree	Mean (SD)
6	Video-based shadowing is effective in improving pronunciation.	44 (77.1%)	10 (17.5%)	3 (5.2%)	0 (0%)	4.75 (0.874)
7	Speaking skills will improve if video-based shadowing improves.	45 (78.9%)	1 (1.7%)	7 (12.2%)	4 (7.0%)	4.74 (0.922)
8	Speaking skills improve with more practice in video-based shadowing.	46 (80.7%)	6 (10.5%)	3 (5.2%)	2 (3.5%)	4.69 (1.064)
9	Pronunciation became better after video-based shadowing.	37 (64.9%)	12 (21.0%)	6 (10.5%)	2 (3.5%)	4.31 (1.091)
10	Don't think video-based shadowing is good for speaking skills.	4 (7.0%)	3 (5.2%)	4 (7.0%)	46 (80.7%)	4.64 (1.355)

Table 3 summarizes the frequency and percentage of responses for the shadowing practice performance questionnaire items of the survey. The results suggest that students perceived consistent progress in their shadowing ability over time. More than three quarters of respondents (73.6%) reported week-to-week improvement in shadowing (Item 11, M=4.61), with an even stronger majority (87.7%) affirming that continued practice leads to improvement (Item 14, M=4.86).

Learners also recognized the importance of practicing at faster speeds (Item 12), with nearly 60% agreeing and an overall mean of 4.25, indicating that many students view speed training as a meaningful component of shadowing. Conversely, the negatively phrased Item 13 received substantial disagreement (57.8%), suggesting that most learners believe that practicing at slightly challenging speeds is beneficial, even if it exceeds their current comfort level.

Table 3: Shadowing Practice Performance

No	Statement	Strongly Agree	Agree	Disagree	Strongly Disagree	Mean (SD)
11	Become better at shadowing week after week	42 (73.6%)	9 (15.7%)	5 (8.7%)	1 (1.7%)	4.61 (1.076)
12	It is important to practice video-based shadowing at a faster speed	34 (59.6%)	17 (29.8%)	4 (7.0)	2 (3.5%)	4.25 (1.105)
13	Not necessary to practice at a speed faster than I can.	6 (10.5%)	7 (12.2%)	11 (19.2%)	33 (57.8%)	3.92 (1.339)
14	Can become better at video-based shadowing if practice more.	50 (87.7%)	6 (10.5%)	1 (1.7%)	0 (0%)	4.86 (0.318)
15	Feedback is very useful to find mistakes.	48 (84.2%)	9 (15.7%)	0 (0%)	0 (0%)	4.84 (0.357)

Table 4 summarizes the frequency and percentage of responses for the satisfaction questionnaire items of the survey. The data indicate that students held highly positive perceptions of video-based shadowing as a meaningful and beneficial learning activity. A strong majority (75.4%) agreed that shadowing practice was a valuable learning experience (Item 16), reflected in a high mean score of 4.81. This positive view is reinforced by the responses to Item 17, where 77.1% of students strongly disagreed that shadowing did not assist their Japanese learning, yielding a mean of 4.75. These results collectively suggest that learners clearly recognized the pedagogical value of shadowing in supporting their language development. Students also perceived cultural benefits from video-based shadowing, with 77.1% strongly agreeing that it enhanced their cultural experience (Item 18).

Table 4: Satisfaction

No	Statement	Strongly Agree	Agree	Disagree	Strongly Disagree	Mean (SD)
16	Shadowing practice is a valuable learning experience.	43 (75.4%)	10 (17.5%)	4 (7.0%)	0 (0%)	4.81 (1.009)
17	Shadowing practice did not assist learning in Japanese.	2 (3.5%)	3 (5.2%)	8 (14.0%)	44 (77.1%)	4.75 (1.156)
18	Shadowing practice using video enhances cultural experience.	44 (77.1%)	9 (15.7%)	2 (3.5%)	2 (3.5%)	4.72 (1.085)

No	Statement	Strongly Agree	Agree	Disagree	Strongly Disagree	Mean (SD)
19	Recommend shadowing practice to friends.	29 (50.8%)	15 (26.3%)	6 (10.5%)	7 (12.2%)	3.67 (1.242)
20	Don't think shadowing improves conversation skills.	2 (3.5%)	8 (14.0%)	9 (15.7%)	38 (66.6%)	4.29 (1.274)

4.2 Participants' responses to open-ended questions

Table 5 summarizes the responses to the open questions of the survey. The qualitative responses highlight several positive aspects of video-based shadowing as perceived by the students. The most frequently mentioned benefit was improvement in speaking ability, reported by 17 respondents. This suggests that learners viewed shadowing as a practical approach to improving oral fluency and overall speaking proficiency.

Listening improvement was the second most cited advantage (15 responses), reinforcing earlier quantitative findings that shadowing helps learners process spoken Japanese more effectively. Pronunciation enhancement (12 responses) and the ability to approximate native-like speech (10 responses) were also frequently noted, suggesting that repeated exposure to authentic input supports more accurate and natural phonological production. In addition to linguistic gains, students identified broader benefits from their learning.

Table 5: Open-Ended Questions Responses (Positive Aspect of Video-Based Shadowing)

No	Positive Aspects	Total Responses
1	Improve speaking	17
2	Improve listening	15
3	Improve pronunciation	12
4	Native-like speech	10
5	Improve vocabulary	9
6	Japanese norm & communication styles	9
7	Improve cultural literacy	6

The responses shown in Table 6 reveal several challenges students encountered during video-based shadowing practice. The most frequently reported difficulty was the fast speed of the input, mentioned by 18 respondents. This suggests that learners often struggled to keep pace with native-level speech, which may have increased cognitive load during practice. Feelings of frustration or stress were the second most common issue (10 responses), indicating that the demanding nature of shadowing can affect learners' emotional comfort, particularly when they perceive the task as overwhelming. Nine respondents reported difficulty in understanding the content, highlighting comprehension barriers that may arise when learners are exposed to rapid and authentic spoken Japanese.

Table 6: Open-Ended Questions Responses (Negative Aspect of Video-Based Shadowing)

No	Negative Aspects	Total Responses
1	Speed too fast	18
2	Frustrated/stressed	10
3	Difficult to understand	9
4	Time consuming	3

Qualitative analysis of the semi-structured interview transcripts showed that the shadowing practice enhances learners' oral skill development. The most frequently cited benefit was improvement in speaking ability, reflecting learners' perception that shadowing directly facilitates oral production. Participants reported that the simultaneous nature of the task pushed them to produce speech under time pressure, mirroring real conversational conditions and thereby building fluency. This aligns with the automatization function of shadowing described by Kadota (2012). Here are some examples of learners' responses:

> "*Practicing shadowing made me more confident in speaking Japanese out loud. I used to be afraid of making mistakes, but repeating what the speaker said helped me get used to the rhythm of Japanese conversation.*"

> "*I learned how to say words more correctly. Listening to the native speaker and immediately trying to copy the sounds made me more aware of my own pronunciation errors.*"

In addition, participants reported that video-based shadowing contributed to vocabulary development, suggesting that the technique fosters incidental lexical acquisition within authentic communicative contexts. Furthermore, participants described gaining awareness of Japanese communicative norms and interactional styles, as illustrated in the following learner responses:

> "*I picked up a lot of new words without trying to memorise them. Because I heard them in real conversations…"*

> "*I began to understand when to use polite or casual Japanese, not just from grammar lessons but from hearing how the speaker in the video naturally switched between styles in different situations.*"

> "*…I noticed how the speaker used certain expressions to be polite or to show they were listening. It made me realise that Japanese communication is about more than just words.*"

In terms of perceived challenges, the most frequently cited challenge was the pace of native-speaker speech in the videos. This finding is theoretically consistent with the cognitive load literature (Shiki et al., 2010; Hamada, 2015): simultaneous listening and speaking places high demands on working memory, and when speech rate exceeds a learner's processing capacity, the dual-task requirement of shadowing becomes cognitively overwhelming rather than beneficial.

> *"The speaker talked so fast that I couldn't keep up. By the time I tried to repeat one sentence, the next one had already passed. It was overwhelming, especially in the first session."*

> *"I think the biggest problem for me was the speed. Japanese native speakers speak very naturally, which is great for learning, but very hard to shadow when you're not used to it."*

Taken together, the themes identified across participant responses reveal a consistent pattern: video-based shadowing within a CLIL framework is perceived by Indonesian JSL learners as a multidimensional learning tool that delivers linguistic, prosodic, lexical, pragmatic, and cultural gains simultaneously, but at a cognitive and affective cost that is non-trivial, particularly for learners encountering natural-rate speech for the first time.

4.3 Integrated Findings: A Convergent Joint Display of Quantitative and Qualitative Strands

Table 7 presents quantitative and qualitative strand integration across eight thematic dimensions derived from the questionnaires, open-ended question responses and analysis of semi-structured interview transcripts. Each row aligns a quantitative finding with its corresponding qualitative theme, a representative participant quotation, and an integrated interpretation that synthesizes both strands into a unified explanatory account. A strand convergence classification is assigned to each dimension to make the nature of the integration explicit and auditable.

Three patterns of convergence were identified. First, convergent dimensions: provide the strongest evidentiary basis for the study's claims, as the finding is not contingent on any single instrument or methodology. Second, expansion dimensions, demonstrate the added analytical value of the mixed-methods approach: specifically, dimensions such as pragmatic register awareness and sociocultural literacy emerged exclusively from open-ended responses, remaining invisible to Likert-scale aggregation. Third, divergent dimensions, yield the study's most theoretically generative findings.

Table 7: Joint Display of Research Findings

Quantitative Finding	Qualitative Theme	Representative Quotation	Integrated Interpretation	Convergence Type
1. Listening Comprehension				
M = 4.69–4.94 >73% agree/strongly agree (Items 1–4) Item 4: 85.9% SA Item 5 (neg.): 77.2% SD	A1b — Positive Listening improvement (n=15, 19.2% of pos. responses)	"Before shadowing, I always got lost when listening to fast Japanese. After practicing, I could catch more words because I was used to how native speakers talk."	Strong agreement across all questionnaire items suggests that learners generally perceived video-based shadowing as beneficial for improving listening comprehension. Participants also reported that	Convergent

		"Participating in shadowing practice through YouTube videos has helped improve my Japanese language proficiency, particularly in pronunciation, listening skills, and learning intonation patterns. Through shadowing, I have developed a better understanding of the rhythm and natural speaking style used by native Japanese speakers." (representative, n=15)	repeated exposure to authentic Japanese speech helped them adapt to natural speech speed and process spoken input more effectively in real time, indicating consistent perceived improvement across their learning experiences.	
2. Speaking Fluency and Oral Production				
M = 4.64–4.75 77–81% agree/strongly agree (Items 6–8) Item 10 (neg.): 80.7% SD	A1a — Positive Speaking improvement (n=17, 21.8% of pos. responses · highest frequency)	"My speaking became smoother after each session. I noticed I could respond faster in class discussions." "Certainly, after engaging in shadowing practice, I gained a better understanding of correct pronunciation, or *hatsuon*." (representative, n=17)	Speaking improvement emerged as the most frequently reported benefit among participants (n=17). Learners described becoming more confident and responding more quickly during classroom interactions, suggesting that the perceived benefits of video-based shadowing extended beyond pronunciation accuracy to include speaking fluency and communicative readiness in authentic contexts.	Convergent
3. Pronunciation Accuracy				
M = 4.31–4.75 Item 6: 77.1% SA	A1c — Positive Pronunciation enhancement	"I learned how to say words more correctly. Listening to the	Although participants strongly perceived shadowing as	Expansion

(pronunciation effective) Item 9: 64.9% SA (lower — individual variation)	(n=12) A2 — Positive Native-like speech (n=10)	native speaker and immediately trying to copy the sounds made me more aware of my own pronunciation errors." (representative, n=12) "The primary benefits I experienced were improvements in listening and pronunciation skills. In addition, I became more familiar with sentence patterns, vocabulary, and intonation commonly used by native speakers, which also contributed to the development of my speaking ability." (representative, n=10)	effective for pronunciation practice, perceptions of personal improvement were slightly more varied. Learners' reflections suggest that some improvements focused on pronunciation accuracy, while others extended to rhythm, intonation, and more natural speech patterns. This indicates that certain learners progressed beyond basic pronunciation practice toward broader phonological awareness, consistent with Kadota's (2012) automatization model.	
4. Shadowing Practice Performance				
M = 3.92–4.86 Item 14: 87.7% SA (practice → improvement) Item 11: 73.6% SA (week-to-week progress) Item 15: 84.2% SA (feedback useful)	Speed too fast (n=18, 45% of neg. responses · highest frequency)	"The speaker talked so fast that I couldn't keep up. By the time I tried to repeat one sentence, the next one had already passed." "When the video is spoken at a fast pace, I still find it difficult to follow. Additionally, newly encountered *kotoba* are often difficult for me to pronounce correctly."	Participants generally viewed shadowing as a skill that improves gradually through repeated practice and feedback. However, many learners also reported difficulty keeping up with the speed of native speakers, indicating that speech processing speed remained a major challenge during the activity. These findings suggest that while learners were motivated to continue practicing,	Divergent

		(representative, n=18)	they still required more structured support, such as adjustable playback speed and step-by-step shadowing practice.	
5. Learner satisfaction and affective response				
M = 4.81 Item 16: 75.4% SA (valuable experience) Item 19: 50.8% SA (recommend to peers) ← lower Item 20 (neg.): 66.6% SD	Frustration/stress (n=10, 25% of neg. responses)	" I found it challenging to follow the pace of native speakers during live conversations, and I often experienced fatigue when participating in prolonged shadowing practice sessions. But I also felt like I had learned something." "Using predetermined videos for shadowing practice that do not match my personal preferences can make the activity more difficult and less engaging. I find that shadowing with anime or dorama that I enjoy is more motivating and encourages me to practice more consistently." (representative, n=10)	Although participants reported high levels of satisfaction with video-based shadowing, several learners also described feelings of fatigue and mental strain during the activities. Many participants perceived the challenge as part of the learning process, suggesting that the demanding nature of shadowing was accompanied by meaningful learning gains. At the same time, some learners appeared more hesitant to recommend the technique for regular use, indicating awareness that the activity requires considerable effort and concentration.	Divergent
6. Cultural and pragmatic competence				
M = 4.72 Item 18: 77.1% SA (cultural experience enhanced) Single item — not decomposed by instrument	A4a — Positive Pragmatic awareness & communication norms (n=9) A4b — Positive Cultural literacy (n=6)	"I began to understand when to use polite or casual Japanese — not just from grammar lessons but from hearing how the speaker	Participants not only perceived cultural benefits from the shadowing activities, but also reported becoming more aware of Japanese communication	Expansion

	Combined: 19.2% of all pos. responses	in the video naturally switched between styles." (representative, n=9) "Video-based shadowing practice greatly helped me understand aspects of Japanese culture and communication. In addition to learning vocabulary and intonation, the videos often provided visual representations of various topics, such as food, festivals, daily life, and Japanese social customs. This made it easier for me to understand conversational contexts and how Japanese is used in real-life situations." (representative, n=6)	styles and everyday social practices. Several learners noticed differences in politeness levels and speech styles used in various situations, while others described gaining a better understanding of daily life and interaction patterns in Japan. These findings suggest that video-based shadowing within a CLIL framework supported not only language development, but also pragmatic and intercultural awareness, consistent with Coyle et al.'s (2010) 4Cs framework.	
7. Cognitive load and comprehension barriers				
M = 3.92 Item 13: 57.8% SD "not necessary to practice at faster speed" — lowest agreement Item 12: 59.6% SA (faster speed important)	B1b — Challenge Difficult to understand content (n=9, 22.5% of neg. responses)	"Sometimes I was just copying sounds without really knowing what was being said. I felt like a parrot. When I didn't understand the meaning, shadowing felt less meaningful and more frustrating." (representative, n=9)	Participants generally recognized the importance of practicing with natural speech speed, yet many still experienced difficulty following fast spoken Japanese during shadowing activities. Learners described situations in which they could imitate sounds and pronunciation patterns but struggled to fully understand the meaning of the	Expansion

			content at the same time. This suggests that rapid shadowing may increase cognitive demands on some learners, indicating the need for additional support during the comprehension phase, such as slower playback speed, repeated listening, or more guided content review within the CLIL-based learning process.	
8. Practical sustainability				
M = 3.67 Item 19: 50.8% SA (recommend to peers) lowest M across all items SD = 1.242 — highest variability across all items	B3 — Challenge Time-consuming (n=3, 7.5% of neg. responses · lowest frequency)	"The activity took a long time because we had to watch the video, then shadow, then understand, then shadow again. It was effective but I'm not sure I can do it every day with all my other assignments." (representative, n=3)	Participants generally viewed video-based shadowing positively; however, some learners appeared more hesitant about recommending the activity for long-term use. Several participants mentioned that the repeated shadowing process required considerable time and concentration, particularly because each session involved multiple stages of listening, comprehension, and repetition. These findings suggest that although the activity was perceived as beneficial for language development, a more flexible or simplified implementation may help improve its practicality and sustainability for continuous learning.	Expansion

5. Discussion

The findings of this study reinforce the relevance of video-based shadowing within the broader framework of Content and Language Integrated Learning (CLIL). As noted by Pires et al. (2024) and Brevik and Moe (2012), CLIL emphasizes the integration of language learning with meaningful content engagement, enabling learners to process language in authentic, context-rich situations. The students' positive perceptions toward shadowing—such as improved listening, pronunciation, speaking fluency, and cultural understanding—reflect how this method supports simultaneous development of linguistic competence and content comprehension.

By shadowing native speakers in various communicative situations, learners engage with real-life discourse, thereby experiencing the type of meaningful learning central to CLIL principles. What is particularly notable, as evidenced by the expansion pattern identified in Dimensions 6 and 7 of the joint display (Table 7), is that these outcomes were not merely reported in Likert-scale items but elaborated upon qualitatively: learners spontaneously described gains in pragmatic register awareness — noticing, for instance, when speakers switched between polite and casual registers — and sociocultural literacy that were not explicitly targeted by instruction. This suggests that the CLIL framework's "4Cs" (Content, Communication, Cognition, and Culture; Coyle et al., 2010) operated synergistically through video-based shadowing, producing cultural and pragmatic outcomes as emergent properties of content-embedded language practice rather than as deliberate instructional goals.

Furthermore, the use of video materials in shadowing aligns with previous findings that highlight the pedagogical value of authentic audiovisual input. Studies by Tahmina (2023) and Oktapiani et al. (2024) emphasize that video-based shadowing exposes learners to natural speech patterns, intonation, and pragmatic cues, which are difficult to replicate through traditional textbook-based instruction. The findings also demonstrate that students perceive YouTube as a highly effective supplementary resource for enhancing their Japanese language proficiency, particularly in listening comprehension, vocabulary acquisition, and cultural and pragmatic understanding (Kristiani, 2021; Balbay & Kilis, 2017).

The strong agreement on these aspects, corroborated by qualitative accounts from 12 participants reporting pronunciation enhancement and 10 reporting approximation of native-like speech (Table 5), suggests that authentic audiovisual input functions on two distinct phonological levels simultaneously: segmental accuracy and suprasegmental prosody. This two-level phonological effect has been similarly observed in shadowing studies conducted in Korean EFL contexts (Hamada, 2019) and Taiwanese Mandarin-speaking EFL learners (Hsieh & Wang, 2013), where authentic speech exposure produced gains in both discrete pronunciation features and broader prosodic naturalness. The consistency of this pattern across typologically distinct L2 contexts —English, Mandarin, and now Japanese in Indonesia — suggests that the mechanism is not language-specific but reflects a generalizable property of the shadowing technique itself when deployed with authentic input: repeated exposure to naturalistic speech activates

phonological noticing (Schmidt, 1990) at multiple levels of processing simultaneously. In this sense, the present findings both replicate and extend the international literature by demonstrating that the dual-level phonological effect is observable in a JSL context where learners operate within a communicative teaching tradition rather than the more form-focused traditions typical of East Asian EFL environments.

The relevance of video-based shadowing becomes even more pronounced in contexts where access to native speakers or immersive environments is limited. As highlighted by Navarrete et al. (2025) and Sari and Margana (2019), such techniques support autonomous learning and provide an alternative pathway for learners in remote or resource-constrained settings to engage with authentic language use. The students' reported improvements, despite challenges such as fast speech or comprehension strain, suggest that video-based shadowing can serve as a flexible and scalable strategy for enhancing language proficiency beyond the physical classroom.

However, the autonomous learning potential of video-based shadowing should not be overstated. The joint display data reveal a critical tension: although 87.7% of participants agreed that additional practice leads to improvement (Item 14, Table 3), the practical sustainability of independent shadowing practice was tempered by reports of time demands (n=3) and the lowest recommendation score across all items (Item 19, M=3.67, SD=1.242). This finding resonates with Al Harrasi's (2023) observation that autonomous digital language learning frequently confronts motivational inconsistency and accountability deficits when learners are left without structured guidance.

In JSL contexts specifically, where learners simultaneously manage multiple courses and limited Japanese exposure outside the classroom, the intensity of multi-session shadowing may reduce long-term sustainability unless integrated into a structured curriculum rather than assigned as supplementary homework. These considerations suggest that while video-based shadowing is theoretically compatible with autonomous learning principles, its implementation design must account for cognitive and logistical demands that can undermine continuity of practice.

A critical engagement with the study's findings requires explicit attention to the negative and potentially unintended outcomes identified through the qualitative strand. The most frequently reported challenge — excessive input speed (n=18, 45% of negative responses, Table 6) — is not merely an implementation inconvenience but reflects a theoretically significant limitation of shadowing with authentic, unmodified native-speaker input. When speech rate exceeds learners' phonological processing capacity, the simultaneous demand of listening and speaking may produce what Shiki et al. (2010) term "phonological echo without semantic engagement" — a state in which learners reproduce sound patterns without constructing meaning, as captured in the qualitative data: participants described feeling "like a parrot" when comprehension failed during rapid shadowing. This outcome is directly contrary to CLIL's foundational principle of

content-driven language learning and represents an unintended consequence of deploying authentic, unscaffolded native-speaker input with intermediate learners. Comparable findings have been reported by Hamada (2015) in Japanese EFL contexts and by Sumiyoshi and Svetanant (2017) in Australian JSL settings, where learners at intermediate proficiency similarly experienced cognitive overload when shadowing natural-rate input without prior comprehension scaffolding. The implication is not that authentic input should be avoided, but that the CLIL-based shadowing procedure used in this study — while effective for most participants — may require a graduated input speed protocol for learners at the lower boundary of intermediate proficiency.

The affective dimension of negative experience also warrants critical interpretation. Ten participants (25% of negative responses) reported frustration and mental exhaustion during shadowing practice. Importantly, the joint display analysis (Dimension 5, Table 7) reveals that this affective strain coexisted with high overall satisfaction (M=4.81, Item 16), producing a divergent pattern that cannot be resolved by either strand alone. The most defensible interpretation, consistent with Vygotsky's (1978) Zone of Proximal Development, is that the emotional cost of shadowing was experienced as proportionate to perceived learning gain, what the educational psychology literature terms "productive struggle" (Kapur, 2016).

However, the risk of this interpretation is that it may normalise affective discomfort without adequately addressing it pedagogically. In practice, instructors implementing video-based shadowing should monitor affective responses across sessions, particularly in learner groups with high foreign language anxiety (Horwitz et al., 1986), as the performative and time-pressured nature of shadowing may disproportionately burden such learners. The absence of a baseline foreign language anxiety measure in the present study represents a potential confound that future research should address.

Several limitations of this study must be acknowledged to contextualize the findings appropriately. First, the reliance on self-reported perceptions rather than objective pre-post-performance measures constrains the interpretive claims that can be made about actual linguistic improvement. Participants' responses reflect perceived development, which, while valuable as a learner-centred outcome measure, may be subject to social desirability bias and the Hawthorne effect (McCarney et al., 2007): awareness of being observed in a pedagogical study may have amplified positive self-reporting, particularly given the absence of anonymous performance-based assessment.

Second, the purposive sampling of university-enrolled, intermediate-level learners from Central Java limits the generalizability of the findings to the broader population of Indonesian JSL learners, who vary considerably in institutional setting, prior exposure to Japanese, and digital literacy. Third, the study did not include a control group receiving conventional instruction without video-based shadowing, which means that observed perceptions of improvement cannot be causally attributed to the intervention itself; they may partly reflect the general

effects of intensive practice regardless of modality. These limitations do not invalidate the findings but establish important boundaries for their interpretation and application.

The findings of this study suggest that video-based shadowing within a CLIL framework constitutes a pedagogically viable and theoretically coherent approach to spoken proficiency development for Indonesian JSL learners at the intermediate level. Its strengths in delivering convergent gains in listening, speaking, pronunciation, and emergent cross-cultural competence are well documented across both quantitative and qualitative strands. Its limitations, such as cognitive overload at natural speech rate, affective strain under time pressure, and sustainability concerns in autonomous settings, are equally important and must be addressed through deliberate instructional scaffolding rather than assumed away. The challenge for practitioners and researchers alike is not whether video-based shadowing works, but how to design its implementation so that its benefits are accessible to a wider range of learners, including those at the boundary of intermediate proficiency and those with higher levels of foreign language anxiety.

6. Conclusion

This study investigated Indonesian JSL learners' perceptions of video-based shadowing within a CLIL framework through a convergent parallel mixed-methods design, confirming that CLIL's 4Cs (Coyle et al., 2010) operated synergistically through content-embedded practice rather than as deliberately instructed outcomes. The joint display analysis identified two convergent, four expansion, and two divergent dimensions, collectively demonstrating that the technique's impact is multidimensional and cannot be fully characterized by aggregate mean scores alone. These findings carry provisional transferability to comparable Southeast Asian JSL contexts where communicative teaching traditions and limited authentic Japanese exposure coexist, though generalizability to beginner or advanced learners and immersive settings cannot be assumed from current data.

7. Implications and Future Directions

The findings carry significant implications for JSL pedagogy, curriculum design, teacher development, and assessment across resource-constrained contexts. Pedagogically, the three-stage shadowing procedure provides a replicable instructional sequence that delivers convergent gains in listening, speaking, and pronunciation while simultaneously producing emergent pragmatic and cultural outcomes; its reliance on freely accessible YouTube content removes the material cost barrier that has historically limited authentic input delivery in Indonesian and comparable Southeast Asian JSL settings. However, the digital divide between urban and rural institutions remains a logistical constraint, and in-class implementation is likely more sustainable than independent homework given the time demands and motivational inconsistency documented in autonomous digital learning contexts (Al Harrasi, 2023). For assessment, oral production tasks embedded within shadowing sessions could capture the phonological and prosodic dimensions of spoken Japanese that paper-based instruments

systematically underassess, while also legitimising shadowing within formal institutional evaluation frameworks. For teacher development, the cognitive, effective, and cultural complexity of managing a CLIL-based shadowing session simultaneously underscores the need for dedicated professional preparation in CLIL methodology, pronunciation pedagogy, and digital tool integration — preparation that current JSL teacher training programs rarely provide.

Beyond the Japanese language context, the theoretical mechanism driving observed gains is not language-specific and could be adapted to other L2 contexts available on YouTube with rich authentic content, including Mandarin, Korean, and Arabic, provided that implementation design accounts for language-specific phonological demands and cultural norms. Regarding sustainability, embedding this approach within a longitudinally sequenced two-semester oral communication curriculum is more likely to produce durable gains than isolated technique adoption.

The rapid evolution of AI-assisted language learning technologies presents new possibilities for addressing the two primary challenges documented in this study: cognitive overload from natural speech rate and affective strain under simultaneous processing demands. Adaptive shadowing platforms capable of adjusting playback speed in real time based on learner performance, combined with AI-powered pronunciation feedback sensitive to Japanese pitch accent and mora timing, could substantially reduce the scaffolding burden currently placed on individual instructors while making the technique more accessible to learners at the lower boundary of intermediate proficiency. Future research should prioritize longitudinal quasi-experimental designs that incorporate pre-post objective performance measures, foreign language anxiety as a baseline covariate, and comparative conditions across learner levels and institutional settings to establish a more robust and differentiated evidence base for the CLIL-shadowing framework documented here.

8. Conflict of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study. The research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

9. Acknowledgments

The authors wish to acknowledge the use of Perplexity, SciSpace, and Grammarly in the writing of this paper. These tools were used to support content development, language editing, and grammar checking. The paper remains an accurate representation of the authors' work and intellectual contributions. This study was fully funded by the Institute for Research and Community Service (LPPM), Universitas Jenderal Soedirman, Riset Dasar Unsoed Grant: 14.302/UN23.34/PT.01/V/2025.

10. References

Adair, J. G. (1984). The Hawthorne effect: A reconsideration of the methodological artifact. *Journal of Applied Psychology, 69*(2), 334–345. https://doi.org/10.1037/0021-9010.69.2.334

Al Harrasi, K. T. S. (2023). Reexamining the online environment for self-directed learning. *Cogent Education*, *10*(2), 1–20. https://doi.org/10.1080/2331186X.2023.2255080

Alkathiri, L. A. (2019). Students' perspectives towards using YouTube in improving EFL learners' motivation to speak. *Journal of Education and Culture Studies, 3*(1), 12. https://doi.org/10.22158/jecs.v3n1p12

Almurashi, W. A. (2016). The effective use of YouTube videos for teaching English language in classrooms as supplementary material at Taibah university in alula. *International Journal of English Language and Linguistics Research, 4*(3), 32-47. https://eajournals.org/wp-content/uploads/The-Effective-Use-of-Youtube-Videos-for-Teaching-English-Language-in-Classrooms-as-Supplementary-Material-at-Taibah-University-in-Alula.pdf

Aprianto, A. (2020). To what extent does YouTube Contents-based Language Learning promote an English proficiency? *Journal of English Language Teaching and Literature, 3*(2), 108-126. https://doi.org/10.47080/jeltl.v3i2.994

Auxier, B. E., & Vitak, J. (2019). Factors motivating customization and echo chamber creation within digital news environments. *Social Media+Society, 5*(2), 1–13. https://doi.org/10.1177/2056305119847506

Balbay, S., & Kilis, S. (2017). Students' perceptions of the use of a YouTube channel specifically designed for an Academic Presentations Skills Course. *Eurasian Journal of Applied Linguistics, 3*(2), 235–251. https://doi.org/10.32601/ejal.461003

Bárcena, E., Read, T., Underwood, J., Obari, H., Cojocnean, D., Koyama, T., Perez-Lora, A., Calle, C., Pompeso, L., Talaván, N., Ávila-Cabrera, J., Ibáñez, A., Vermeulen, A., Jordan, M., Arús-Hita, J., Rodríguez, P., Castrillo, M. D., Kiévi, J., & Kukulska-Hulme, A. (2015). *State of the art of language learning design using mobile technology: sample apps and some critical reflection*. In J. Obari, H. Cojocnean, D. Koyama, T. Perez-Lora, A. Calle, C. Pompeso, L. Talaván, N. Ávila-Cabrera, J. Ibáñez, A. Vermeulen, A. Jordan, J. Arús-Hita, J. Rodríguez, P. Castrillo, M. D. Kiévi, & J. Kukulska-Hulme (Eds.), *Critical CALL: Proceedings of the EUROCALL 2015 Conference, Padova, Italy* (pp. 36–43). https://doi.org/10.14705/rpnet.2015.000307

Bovee, N., & Stewart, J. (2008). The utility of shadowing. In A. Stoke (Ed.), *JALT 2008 Conference Proceedings* (pp. 888–900). Japan Association for Language Teaching.

Brevik, M., & Moe, E. (2012). Effects of CLIL teaching on language outcomes. In D. Tsagari & I. Csépes (Eds.), *Collaboration in language testing and assessment* (pp. 213–227). Peter Lang. https://doi.org/10.3726/978-3-653-01526-3/16

Brinton, D. M. (2001). *The use of media in language teaching*. The British Council.

Byun, U., Jang, M., & Baek, H. (2023). The effect of YouTube comment interaction on video engagement: focusing on interactivity centralization and creators' interactivity. *Online Information Review, 47*(6), 1083–1097. https://doi.org/10.1108/OIR-04-2022-0217

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: Complex or simple? Research case examples. *Journal of Research in Nursing, 25*(8), 652–661. https://doi.org/10.1177/1744987120927206

Chhabra, P. (2012). Use of e-learning tools in teaching English. *International Journal of Computing & Business Research*. Proceedings of *I-Society 2012*, Guru Kashi University, Talwandi Sabo Bathinda, Punjab, India. Retrieved from http://www.researchmanuscripts.com/isociety2012/9.pdf

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.

Daeli, R. (2024). The effectiveness of language-related YouTube videos on EFL learning motivation: A survey of young adult Indonesian learners' perceptions. *Working Papers in Language Pedagogy, 18*, 37–55. https://doi.org/10.61425/wplp.2023.18.37.55

Dalton-Puffer, C. (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of Applied Linguistics*, 31, 182–204. https://doi.org/10.1017/S0267190511000092

Danial, M. (2022). YouTube-sourced videos as teaching media for listening comprehension: An optimizing of authentic-updated learning source. *English Language, Linguistics, and Culture International Journal, 2*(3), 196–203. https://doi.org/10.24252/elstic-ij.v2i3.33464

Derwing, T. M. (2010). Utopian goals for pronunciation teaching. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 1st Pronunciation in Second Language Learning and Teaching Conference* (pp. 24–37). Iowa State University.

Fitri, D., & Ma'rifah, U. (2022). The use of subtitled video animation to improve students' vocabulary at Ban Erawan School Thailand. *English Education Journal, 12*(4), 526–537. https://doi.org/10.15294/eej.v12i4.66626

Hamada, Y. (2011). Psychological Aspects of Shadowing Training. *The Japan Association for Developmental Education, 6*(2), 60–71.

Hamada, Y. (2015). Uncovering shadowing as an EFL teaching technique for listening: Learners' perceptions, self-confidence, and motivation. *Akita University Annual Research Report on General Education, 9*(22), 9–22.

Hamada, Y. (2016). Shadowing: Who benefits and how? Uncovering a booming EFL technique for listening comprehension. *Language Teaching Research, 20*(1), 35–52. https://doi.org/10.1177/1362168815597504

Hamada, Y. (2019). Shadowing: What is it? How to use it. Where will it go? *RELC Journal, 50*(3), 386–393. https://doi.org/10.1177/0033688218771380

Hamada, Y., & Suzuki, Y. (2022). Situation shadowing in a framework of deliberate practice: A guide to using 16 techniques. *RELC Journal 55*(1), 1–9. https://doi.org/10.1177/00336882221087508

Haryono, E. P. (2019). CLT in ESL and EFL Context: A Challenge to Civil Engineering Teachers in Indonesia. *Jurnal Teknik Sipil Terapan*, *1*(2), 45–50. https://doi.org/10.47600/JTST.V1I2.235

Hia, M. (2021). The effect of using YouTube on students' motivation in learning English vocabulary. *Dialectical Literature and Education Journal, 6*(2), 62–71. https://doi.org/10.51714/dlejpancasakti.v6i2.61.pp.62-71

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125–132. https://doi.org/10.1111/j.1540-4781.1986.tb05256.x

Hsieh, K., Dong, D., & Wang, Y. (2013). A preliminary study of applying shadowing technique to English intonation instruction. *Taiwan Journal of Linguistics, 11*(2), 43–66.

Hughey, J. (2020). Individual personalized learning. *Educational Considerations, 46*(2), 1–8. https://doi.org/10.4148/0146-9282.2237

Jin, Z., & Webb, S. (2021). Does writing words in notes contribute to vocabulary learning? *Language Teaching Research, 29*(2), 1–40. https://doi.org/10.1177/13621688211062184

Kadota, S. (2012). *Shadowing, ondoku to eigo shutoku no kagaku* [The science of shadowing, oral reading and English acquisition]. CosmoPier Publishing Company.

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kato, S. (2009). Kokueigo noryoku shomei shutoku wo mezashita listening shido no kosatsu [Listening activities for the acquisition of aviation English proficiency test]. *Bulletin of Chiba University Language and Culture, 3*, 47–59.

Katz-Buonincontro, J. (2024). Convergent mixed methods designs. In J. Katz-Buonincontro, *How to mix methods: A guide to sequential, convergent, and experimental research designs* (pp. 73–82). American Psychological Association. https://doi.org/10.1037/0000404-005

Khan, A., Shar, A., & Kashif, M. (2024). Using YouTube for achieving learning efficiency: A case study of Karachi University students. *SBBU Journal of Social Sciences, 1*(1), 47–59.

Kristiani, P. E., & Pradnyadewi, D. A. M. (2021). The effectiveness of YouTube as learning media in improving learners' speaking skills. *The Art of Teaching English as a Foreign Language, 2*(1), 7–11. https://doi.org/10.36663/tatefl.v2i1.97

Kurata, K. (2007). A fundamental study on the cognitive mechanism of shadowing in Japanese: From the viewpoint of oral reproduction, memory span and sentence structure. *Bulletin of the Graduate School of Education, Hiroshima University, 56*(2), 259–265.

Lambert, S. (1992). Shadowing. *Meta*, 37 (2), 263–273. https://doi.org/10.7202/003378ar

Levis, J. (2020). Revisiting the intelligibility and nativeness principles. *Journal of Second Language Pronunciation, 6*(3), 310–328.

Lewis, T. (2014). Learner Autonomy and the Theory of Sociality. In: Murray, G. (eds) Social Dimensions of Autonomy in Language Learning. Palgrave Macmillan, London. https://doi.org/10.1057/9781137290243_3

Lichtman, K., & VanPatten, B. (2021). Was Krashen right? Forty years later. *Foreign Language Annals, 54*(2), 283–305. https://doi.org/10.1111/flan.12552

Long, M. H. (2015). *Second language acquisition and task-based language teaching*. Wiley-Blackwell.

McCarney, R., Warner, J., Iliffe, S., van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: A randomised, controlled trial. *BMC Medical Research Methodology, 7*, Article 30. https://doi.org/10.1186/1471-2288-7-30

Mochizuki, M. (2006). On the application of shadowing method to Japanese language education: With special reference to relevance of learners' Japanese competence to their self-assessment of the effects of shadowing. *Kansai University Audio-Visual Education, 29*, 37–53.

Murphy, J. (2014). Myth 7: Teacher training programs provide adequate preparation in how to teach pronunciation. In L. Grant (Ed.), *Pronunciation myths: Applying second language research to classroom teaching* (pp. 188–224). University of Michigan Press. https://doi.org/10.3998/mpub.4584330

Mutiarani, M., & Rusiana, A. (2021). Stimulating students speaking using english speeches youtube channel. *JALL (Journal of Applied Linguistics and Literacy), 5*(2), 40 - 55. http://dx.doi.org/10.25157/jall.v5i2.5106

Nasution, A. K. R. (2019). YouTube as a media in English language teaching (ELT) context: Teaching procedure text. *Utamax, 1*(1), 29–33. https://doi.org/10.31849/utamax.v1i1.2788

Navarrete, E., Nehring, A., Schanze, S., Ewerth, R., & Hoppe, A. (2025). A closer look into recent video-based learning research: A comprehensive review of video

characteristics, tools, technologies, and learning effectiveness. *International Journal of Artificial Intelligence in Education, 35*(4), 1631-1694. https://doi.org/10.1007/s40593-025-00481-x

Newton, J. M., & Nation, I. S. P. (2021). *Teaching ESL/EFL listening and speaking* (2nd ed.). Routledge. https://doi.org/10.4324/9780429203114

Nofrika, I. (2019). EFL students' voices: The Role of YouTube in Developing English Competences. *Journal of Foreign Language Teaching and Learning, 4*(1). https://doi.org/10.18196/ftl.4138

Oktapiani, I. U., Ilham, I., Humaira, H., Bafadal, F., Adebayo, M., & Fasih, P. (2024). The effectiveness of English-subtitled videos in enhancing speaking, listening and vocabulary skills: A meta-analysis. *Journal of Languages and Language Teaching, 12*(1), 56. https://doi.org/10.33394/jollt.v12i1.9501

Omar, H., & Umehara, M. (2010). Using a "shadowing technique" to improve English pronunciation of deficient adult Japanese learners: Action research on expatriate Japanese adult learners. *The Journal of Asia TEFL*, *7*(2): 199–230.

Pires, L., Manchebon, M. J., Mediavilla, M., & Sancho, J. M. G. (2024). New evidence on the impact of learning in a foreign language on educational outcomes. *Studies in Educational Evaluation, 83*, Article 101386, 1–11. https://doi.org/10.1016/j.stueduc.2024.101386

Putri, H. R. (2022). The effectiveness of teaching speaking by using drilling method through YouTube. *Journal of English Teaching, Literature, and Applied Linguistics, 6*(1), 70–81. https://doi.org/10.30587/jetlal.v6i2.4069

Putri, H. F. H., Wijayanto, A., & Supriyadi, S. (2020). Strengths and weaknesses of self-regulated learning through YouTube: Indonesian EFL students' perceptions. *ELS Journal on Interdisciplinary Studies in Humanities, 3*(4), 531–542. https://doi.org/10.34050/elsjish.v3i4.11749

Qomariyah, S. S., Permana, D., & Hidayatullah, H. (2021). The effect of youtube video on students' listening comprehension performance. *Jo-ELT (Journal of English Language Teaching) Fakultas Pendidikan Bahasa & Seni Prodi Pendidikan Bahasa Inggris IKIP, 8*(1), 67. https://doi.org/10.33394/jo-elt.v8i1.3837

Renandya, W. A., & Farrell, T. S. C. (2011). 'Teacher, the tape is too fast!' Extensive listening in ELT. *ELT Journal, 65*(1), 52–59. https://doi.org/10.1093/elt/ccq015

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

Sari, Y. M., Margana. (2019). YouTube as a learning media to improve the student's speaking ability in 21st century. *JELTL (Journal of English Language Teaching and Linguistics), 4*(2), 263–273. https://doi.org/10.21462/jeltl.v4i2.296

Satriani, E., Wahyuni, S., Etfita, F., Ahmad, A., & Novitri, S. (2023). Autonomous learning material with YouTube: Case study for intensive reading course. *AL-ISHLAH: Jurnal Pendidikan, 15*(2), 2562–2569. https://doi.org/10.35445/alishlah.v15i2.3500

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Shanmugasundaram, M., & Tamilarasu, A. (2023). The impact of digital technology, social media, and artificial intelligence on cognitive functions: A review. *Frontiers in Cognition, 2*, 1–11. https://doi.org/10.3389/fcogn.2023.1203077

Shoufan, A., & Mohamed, F. (2022). YouTube and education: A scoping review. *IEEE Access, 10*, 125576–125599. https://doi.org/10.1109/ACCESS.2022.3225419

Shiki, O., Mori, Y., Kadota, S., & Yoshida, S. (2010). Exploring differences between shadowing and repeating practices: An analysis of reproduction rate and types of

reproduced words. *Japan Society of English Language Education, 21*, 81–90. https://doi.org/10.20581/arele.21.0_81

Sianna, Ramlah, & Salasiah. (2018). Teaching writing with authentic video in EFL classroom. *ELS Journal on Interdisciplinary Studies in Humanities, 1*(4), 448–458. https://doi.org/10.34050/els-jish.v1i4.5395

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Spiller-Bosatra, E., Darò, V., Fabbro, F., & Bosatra, A. (1990). Audio-phonological and neuropsychological aspects of simultaneous interpretation role of auditory shadowing. *Scandinavian Audiology*, *19*(2), 81–87. https://doi.org/10.3109/01050399009070757

Sumiyoshi, H., Svetanant, C. (2017). Motivation and attitude towards shadowing: Learners' perspectives in Japanese as a foreign language. *Asian-Pacific Journal of Second and Foreign Language Education*, 2(16), 1-21. https://doi.org/10.1186/s40862-017-0039-6

Tahmina, T. (2023). Students' Perception of the Use of YouTube in English Language Learning. *Journal of Languages and Language Teaching, 11*(1), 151. https://doi.org/10.33394/jollt.v11i1.6883

Tamai, K. (1992). "Follow-up No Chokairyoku Kojo ni Oyobosu Koka Oyobi "Follow-up" Noryoku to Chokairyoku No Kankei. dai 4 kai "Eiken" Kenkyu Josei Hokoku [The Effect of Follow-up on Listening Comprehension]." *STEP Bulletin* 4:48–62.

Tamai, K. (1997). "Shadowing No Koka to Chokai Process ni Okeru Ichizuke [The Effectiveness of Shadowing and Its Position in the Listening Process]." *Current English Studies* 36:105–116.

Van, L. K., Dang, T. A., Pham, D. B. T., Vo, T. T. N., & Pham, V. P. H. (2021). The effectiveness of using technology in learning English. *AsiaCALL Online Journal, 12*(2), 24–40. https://asiacall.info/acoj/index.php/journal/article/view/26

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Whitworth, B., & Rose, H. (2025). A systematic review of research on the use of shadowing for second language pronunciation teaching. *Research Synthesis in Applied Linguistics, 1*(2), 239–269. https://doi.org/10.1080/29984475.2025.2546827

Yates, L. (2011). Language, interaction and social inclusion in early settlement. *International Journal of Bilingual Education and Bilingualism, 14*(4), 457–471.

Yuyun, I., & Simamora, F. Y. (2021). The use of YouTube to support EFL students' listening skills. *ELLTER Journal, 2*(2), 1–12. https://doi.org/10.22236/ellter.v2i2.7512

Zeng, S. (2020). The potential of online technology for language learning. *English Language Teaching, 13(*10), 23–37. https://doi.org/10.5539/elt.v13n10p23

Zengin, O., & Onder, M. E. (2021). Educational quality of YouTube videos on musculoskeletal ultrasound. *Clinical Rheumatology, 40*(10), 4243–4251. https://doi.org/10.1007/s10067-021-05793-6

Zielinski, B. (2011). The social impact of pronunciation difficulties: Confidence and willingness to speak. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 23–38). Iowa State University.

Zhang, H., & Lee, J. (2023). Exploring audience engagement in YouTube vlogs through consumer engagement theory: The case of U.K. beauty vlogger Zoe Sugg. *First Monday, 28*(4), 1–14. https://doi.org/10.5210/fm.v28i4.11822

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics, 13*(1), 75–98. https://doi.org/10.1007/s10683-009-9230-z

Zoss, J. (2016). What do adult English learners say about their pronunciation and linguistic self-confidence? *MinneTESOL Journal, 32*(2), 1–14. https://minnetesoljournal.org/wp-content/uploads/2018/10/ZossFall2016MinneTESOLJournal.pdf.

Zulkflee, Z., Nor Azmi, N., Sheikh Kamaruzaman, S. S., Prakas, J., Saputra, S., & Maniam, M. (2022). Students' perceptions in using YouTube to improve their speaking skill: A cross-national study of Malaysia and Indonesia. *International Journal of Academic Research in Business and Social Sciences, 12*(10), 3240–3260. https://doi.org/10.6007/IJARBSS/v12-i10/15378

Appendix 1: Questionnaire instruments

Judul Penelitian: *Video-Based Shadowing in a Content and Language Integrated Learning (CLIL) Framework: Study of Indonesian Japanese as Second Language (JSL) Learners' Perceived Listening, Pronunciation, and Speaking Development*

Petunjuk Pengisian: Bacalah setiap pernyataan di bawah ini dengan cermat. Berikan tanda centang (✓) pada kolom yang paling sesuai dengan pengalaman dan pendapat Anda selama mengikuti sesi latihan shadowing berbasis video YouTube. Tidak ada jawaban benar atau salah. Jawaban Anda bersifat rahasia dan hanya digunakan untuk keperluan penelitian.

Skala penilaian yang digunakan:

Skor	**1**	**2**	**3**	**4**	**5**
Arti	Sangat Tidak Setuju (STS)	Tidak Setuju (TS)	Netral (N)	Setuju (S)	Sangat Setuju (SS)

No.	**Pernyataan**	**SS (5)**	**S (4)**	**N (3)**	**TS (2)**	**STS (1)**
A. Kompetensi Menyimak						
1	Shadowing berbasis video efektif dalam meningkatkan kemampuan menyimak saya.					
2	Kemampuan menyimak saya meningkat seiring dengan peningkatan kemampuan shadowing berbasis video saya.					
3	Kemampuan menyimak saya meningkat dengan semakin banyak berlatih shadowing berbasis video.					
4	Saya menjadi lebih baik dalam menyimak setelah berlatih shadowing berbasis video.					
5	Saya tidak merasa shadowing berbasis video bermanfaat untuk kemampuan menyimak saya. (*)					
B. Kompetensi Berbicara						
6	Shadowing berbasis video efektif dalam meningkatkan kemampuan pengucapan/pelafalan saya.					
7	Kemampuan berbicara saya akan meningkat jika kemampuan shadowing berbasis video saya meningkat.					
8	Kemampuan berbicara saya meningkat dengan semakin banyak berlatih shadowing berbasis video.					
9	Pengucapan saya menjadi lebih baik setelah melakukan shadowing berbasis video.					
10	Saya tidak merasa shadowing berbasis video bermanfaat untuk kemampuan berbicara saya. (*)					
C. Performa dalam Sesi Latihan Shadowing						
11	Saya semakin baik dalam melakukan shadowing dari minggu ke minggu.					

No.	Pernyataan	SS (5)	S (4)	N (3)	TS (2)	STS (1)
12	Penting untuk berlatih shadowing berbasis video dengan kecepatan yang lebih tinggi.					
13	Saya tidak perlu berlatih dengan kecepatan yang melampaui kemampuan saya saat ini. (*)					
14	Saya bisa menjadi lebih baik dalam shadowing berbasis video jika saya berlatih lebih banyak.					
15	Umpan balik (feedback) sangat berguna untuk menemukan kesalahan dalam praktik shadowing saya.					
D. Kepuasan terhadap Sesi Latihan Shadowing						
16	Praktik shadowing merupakan pengalaman belajar yang berharga.					
17	Praktik shadowing tidak membantu saya dalam belajar bahasa Jepang. (*)					
18	Praktik shadowing menggunakan video meningkatkan pemahaman budaya saya.					
19	Saya akan merekomendasikan praktik shadowing kepada teman-teman saya.					
20	Saya tidak merasa shadowing meningkatkan kemampuan percakapan saya. (*)					

NB: Item yang diberi tanda (*) merupakan pernyataan negatif.

E. Pertanyaan Terbuka

Jawablah pertanyaan-pertanyaan berikut secara singkat dan jujur berdasarkan pengalaman Anda selama mengikuti sesi latihan shadowing berbasis video.

No.	Pertanyaan
1.	Apa manfaat utama yang Anda rasakan dari latihan shadowing berbasis video YouTube dalam pembelajaran bahasa Jepang Anda? Jelaskan secara singkat. Jawaban:
2.	Apa tantangan atau kesulitan utama yang Anda hadapi selama mengikuti latihan shadowing berbasis video YouTube? Jelaskan secara singkat. Jawaban:
3.	Apakah latihan shadowing berbasis video membantu Anda memahami aspek budaya dan komunikasi bahasa Jepang? Jika ya, berikan contoh konkret. Jawaban:
4.	Apakah Anda merasa latihan shadowing ini dapat Anda lanjutkan secara mandiri di luar kelas? Apa yang mendukung atau menghambat keberlanjutan latihan tersebut? Jawaban:

Appendix 2: Semi-Structured Interview Guide

Judul Penelitian: *Video-Based Shadowing in a Content and Language Integrated Learning (CLIL) Framework: Study of Indonesian Japanese as Second Language (JSL) Learners' Perceived Listening, Pronunciation, and Speaking Development*

Tujuan Wawancara	Menggali secara mendalam persepsi, pengalaman, dan pandangan peserta tentang manfaat dan tantangan praktik shadowing berbasis video YouTube dalam konteks pembelajaran CLIL bahasa Jepang.
Durasi	45–60 menit per sesi
Format	Semi-terstruktur; pewawancara dapat mengajukan pertanyaan lanjutan (probing) sesuai respons peserta
Bahasa	Bahasa Indonesia (atau campuran Bahasa Indonesia dan Bahasa Jepang sesuai kenyamanan peserta)
Rekaman	Dengan persetujuan peserta, wawancara direkam secara audio untuk keperluan transkripsi. Data digunakan secara anonim.

Daftar Pertanyaan Wawancara

No.	Pertanyaan Wawancara
Tema 1: Persepsi Umum tentang Shadowing Berbasis Video	
1	Bagaimana kesan pertama Anda ketika pertama kali melakukan latihan shadowing menggunakan video YouTube bahasa Jepang? Ceritakan pengalaman tersebut.
2	Secara keseluruhan, bagaimana Anda menilai efektivitas latihan shadowing berbasis video dalam membantu Anda belajar bahasa Jepang? Mengapa demikian?
3	Dibandingkan dengan metode belajar bahasa Jepang lainnya yang pernah Anda gunakan, apa yang membuat shadowing berbasis video berbeda atau lebih baik/kurang efektif menurut Anda?
Tema 2: Peningkatan Kompetensi Linguistik	
4	Apakah Anda merasakan peningkatan dalam kemampuan menyimak setelah mengikuti latihan shadowing? Jika ya, pada aspek apa yang paling terasa perubahannya? Berikan contoh konkret.
5	Bagaimana pengaruh latihan shadowing terhadap pengucapan/pelafalan Anda dalam bahasa Jepang? Apakah ada aspek pelafalan atau intonasi tertentu yang mengalami perubahan?
6	Apakah Anda merasakan perbedaan dalam kelancaran berbicara setelah rutin melakukan shadowing? Ceritakan pengalaman Anda.
7	Sejauh mana latihan shadowing membantu Anda mengenali dan meniru pola intonasi, ritme, dan tempo bicara penutur asli bahasa Jepang?
Tema 3: Kesadaran Budaya dan Kompetensi Pragmatik	
8	Apakah Anda memperhatikan aspek budaya Jepang atau gaya komunikasi tertentu melalui video yang digunakan dalam latihan shadowing? Apa yang Anda pelajari dari situ?
9	Apakah Anda menjadi lebih sadar terhadap perbedaan penggunaan bahasa (misalnya penggunaan bahasa formal/keigo vs. bahasa santai/casual) setelah berlatih shadowing dengan video? Jelaskan.

No.	Pertanyaan Wawancara
10	Menurut Anda, apakah video dari channel 'Speak Japanese Naturally' yang digunakan dalam latihan ini cukup mencerminkan situasi komunikasi sehari-hari di Jepang? Apa kelebihan dan kekurangannya?
Tema 4: Tantangan Kognitif dan Respons Afektif	
11	Apa tantangan terbesar yang Anda hadapi selama latihan shadowing? Bagaimana Anda mengatasinya?
12	Apakah Anda pernah merasa frustasi, kewalahan, atau lelah secara mental selama sesi shadowing? Pada kondisi seperti apa hal itu terjadi?
13	Bagaimana perasaan Anda ketika berhasil menyelesaikan satu sesi shadowing dengan baik? Apakah hal tersebut memengaruhi motivasi belajar Anda?
14	Apakah kecepatan bicara penutur asli dalam video menjadi hambatan bagi Anda? Bagaimana hal itu memengaruhi proses belajar Anda?
Tema 5: Keberlanjutan dan Rekomendasi	
15	Apakah Anda merasa latihan shadowing berbasis video YouTube ini dapat dilanjutkan secara mandiri di luar kelas? Apa faktor-faktor yang mendukung atau menghambat keberlanjutannya?
16	Jika Anda harus merekomendasikan (atau tidak merekomendasikan) teknik ini kepada sesama mahasiswa bahasa Jepang, apa yang akan Anda katakan? Mengapa?
17	Apakah ada saran atau perbaikan yang ingin Anda usulkan terkait cara pelaksanaan latihan shadowing ini agar lebih efektif dan berkelanjutan?