

Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos

¹Prayudha*, ¹Rahmi Munfangati, ¹Fitri Heyza Galuh Kirana

*¹Universitas Ahmad Dahlan, Jl. Ringroad Selatan, Tamanan, Banguntapan, Bantul,
Yogyakarta, 55191, Indonesia*

***Corresponding Author**

Email: prayudha@uad.ac.id

Received:
01 May 2026

Revised:
10 July 2026

Accepted:
16 July 2026

Published:
27 July 2026

Abstract

This study investigates how cohesion is realized in English project videos produced by communication studies students within an English for Specific Purposes (ESP) context. While cohesion has traditionally been examined as a structural feature of written discourse, limited attention has been given to its role in spoken and multimodal communication. Adopting a qualitative discourse-analytic approach grounded in Halliday and Hasan's framework, this study analyzes 13 student-generated self-introduction videos to identify patterns of grammatical and lexical cohesion and their functional roles in meaning-making. The analysis identified 90 cohesive instances across five dominant categories: personal reference, conjunction, lexical repetition, demonstrative reference, and collocation. The findings reveal that personal reference serves as the primary organizing mechanism, supporting identity construction and audience alignment, while lexical repetition functions as a strategic resource for thematic continuity. Conjunctions provide minimalist yet effective logical structuring, and demonstrative references frequently operate across both verbal (spoken language) and visual (on-screen gesture and imagery) channels, indicating the presence of multimodal cohesion. These patterns suggest that students employ cohesion not merely as a structural linguistic system but as an adaptive communicative strategy shaped by real-time production demands and digital discourse contexts. Based on this limited dataset, this study offers a preliminary contribution toward reframing cohesion as a functional, strategic, and multimodal resource rather than a purely structural feature, warranting further investigation with larger corpora. Pedagogically, the findings highlight the importance of integrating discourse-based and multimodal approaches in ESP instruction to support students' communicative competence in digital environments.

Keywords: Cohesion; communicative competence; ESP; multimodal discourse analysis

INTRODUCTION

The field of Applied Linguistics has increasingly emphasized the role of language in digitally mediated communication, where meaning is constructed through multimodal resources rather than purely textual forms (Bruce, 2018; Yu &

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

Lowie, 2020). Recent global trends show that digital platforms such as video-sharing media and social networks have transformed how individuals produce and interpret discourse, particularly in educational settings (Barrot, 2022; Hafner & Ho, 2020). In higher education, especially in communication-related disciplines, students are expected to demonstrate not only linguistic accuracy but also the ability to construct coherent and engaging messages across multimodal environments (Jiang & Luk, 2016; Wigham & Satar, 2021).

This shift reflects a broader pedagogical transition toward performance-based and authentic assessment, where language use is evaluated in real-world communicative tasks rather than controlled written exercises (Barkaoui, 2026). Consequently, discourse competence, particularly cohesion and coherence, has become a central concern in understanding how learners effectively communicate in digital spaces (H. Liu et al., 2024; Lorenzo-Dus et al., 2016).

Previous studies on discourse analysis have extensively examined cohesion and coherence in various contexts, yet most have focused on written texts rather than spoken digital discourse. For instance, Prayudha (2016) demonstrated that grammatical cohesion (e.g., conjunctions) and lexical cohesion (e.g., repetition) are essential for maintaining textual clarity and unity, a principle that remains critical in AI-assisted and multimodal digital writing where learners must construct coherent meaning across multiple modes. Human learners outperform AI in deep cohesion within narrative writing, emphasizing the importance of semantic connectivity (Zhou et al., 2023). Meanwhile, Pamuji et al. (2025) explored video-based ESP learning and found that digital projects enhance students' communicative competence through authentic language use. Despite these contributions, discourse-analytic studies of cohesion in student-produced video genres remain narrow in scope. Lindenberg (2023), for instance, examined intersemiotic cohesion in university students' online presentations, showing how spoken commentary and slide visuals were linked through cohesive devices. However, this work centers on formal academic presentations rather than the more performative, identity-oriented genre of self-introduction or promotional project videos. Studies explicitly analyzing discourse-level cohesion in such videos, particularly within communication studies department, where students must balance linguistic accuracy with audience-facing persuasive and identity-construction goals, remain limited.

This gap reveals a critical limitation in current research within Applied Linguistics, where cohesion is still predominantly conceptualized within written discourse frameworks. Existing studies tend to evaluate cohesion based on frequency counts or grammatical accuracy, often neglecting its functional role in constructing meaning within multimodal communication (McNamara et al., 2014). Moreover, there is insufficient attention to how students deploy cohesive devices strategically in real-time spoken discourse, particularly in video-based academic tasks that integrate visual, auditory, and linguistic elements (Jewitt et al., 2016). In English for Specific Purposes (ESP) contexts, this gap is even more pronounced, as learners are required to adapt language use to professional communication demands, yet research rarely examines how cohesion supports identity construction, persuasion, and audience engagement in such settings (Alfaridhi et al., 2026; Prayudha et al., 2026). Therefore, a more

nuanced and context-sensitive analysis is needed to understand how cohesion functions beyond textual linkage, particularly in digital and performance-based discourse.

Addressing this gap, the present study contributes to Applied Linguistics by extending the application of cohesion theory to a new context: multimodal spoken discourse in student-generated English project videos. Rather than proposing a new theoretical framework, this study offers empirical support for viewing cohesion as a strategic resource that operates across both linguistic and visual dimensions (Jones et al., 2015), rather than merely a structural linguistic feature. Unlike previous research that focuses on written texts or isolated linguistic units, this study analyzes English project videos as authentic communicative performances. The contribution of this research lies in three aspects. First, it applies classical cohesion theory proposed by Halliday and Hasan (2014) to multimodal contexts, demonstrating how cohesive devices function in spoken video discourse. Second, it connects discourse analysis with ESP pedagogy, highlighting how cohesion contributes to professional communication skills in communication studies students. Third, it provides an empirical examination of student-generated digital content, offering insights into how learners construct meaning, organize ideas, and engage audiences in real-world communication scenarios. This perspective illustrates that cohesion functions not only as a linguistic mechanism but also as a communicative strategy within digital discourse.

Based on this framework, the present study aims to investigate the use of cohesion devices in English project videos produced by communication studies students. Specifically, the study seeks to identify the types of grammatical and lexical cohesion devices employed, analyze how these devices contribute to coherence and clarity in spoken discourse, and explore their pedagogical implications for ESP instruction. Accordingly, the research addresses the following questions: (1) What cohesion devices are found in English project videos of communication studies students? (2) How do the cohesion devices contribute to discourse coherence and communicative effectiveness? By answering these questions, the study aims to provide both theoretical and practical contributions to Applied Linguistics and ESP, particularly in understanding discourse competence in digital and multimodal communication contexts.

Review of Literature

Discourse Analysis in Multimodal Contexts

Within Applied Linguistics, discourse analysis has evolved from a focus on linguistic structures toward a broader understanding of language as social practice embedded in cultural, ideological, and multimodal contexts (Flowerdew, 2012; Gee, 2014). Contemporary research emphasizes that discourse is not merely a sequence of sentences but a form of meaning-making shaped by context, power relations, and communicative purposes. Digital discourse studies extend these macro-level concerns to multimodal platforms, showing that communication increasingly occurs through the interaction of verbal, visual, and auditory elements (D'Angelo & Maci, 2021). This extension,

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

however, has largely retained the same emphasis on discourse as social practice rather than on the specific cohesive resources speakers deploy in real time.

The theoretical contributions of van Dijk (2014) and Fairclough (2015) remain influential in this shift, though they diverge in emphasis. Van Dijk's socio-cognitive model foregrounds how discourse is mediated by shared mental representations and social cognition, while Fairclough's three-dimensional framework prioritizes how discourse enacts and contests power relations within sociocultural structures. Both approaches, however, treat discourse primarily as a social or ideological phenomenon, and they offer limited tools for examining the internal linguistic mechanisms, such as cohesion, that hold multimodal texts together at the level of language use. This gap between macro-social theories of discourse and the micro-linguistic analysis of cohesion motivates the more text-internal, Hallidayan approach adopted in this study.

Cohesion as a Functional and Strategic Resource

Cohesion remains a central concept in discourse analysis, traditionally defined by Halliday and Hasan as the set of linguistic resources that create textual unity through grammatical and lexical relations. This structural view has since been challenged by scholars such as (Hyland, 2019), who argues that cohesion should not be treated solely as a formal property of text but as a functional and strategic resource in communication, in which devices such as reference, conjunction, and lexical repetition serve interpretive and interactional purposes rather than acting merely as connective tissue. These two positions are not simply sequential refinements of one another; they reflect a genuine tension in the field.

Crossley et al. (2016) reaffirm the structural view empirically, demonstrating that lexical cohesion, particularly repetition and collocation, occurs with high frequency in second language learner discourse. McNamara et al. (2014) complicate this picture, however, showing through computational discourse analysis that the mere presence of cohesive markers does not guarantee coherence, since effective discourse also depends on semantic integration and contextual appropriateness. Read together, these two studies suggest that frequency-based accounts of cohesion, however useful for description, cannot alone explain how cohesive devices contribute to meaning-making, a limitation this study addresses by analyzing not only the distribution of cohesive devices but their functional role in each instance.

Cohesion in Multimodal and Digital Discourse

Recent scholarship has begun to address the role of cohesion in multimodal communication, particularly in video-based discourse. Jewitt et al. (2016) argue that meaning is distributed across different modes, including language, image, gesture, and sound, that interact to create coherent communication, implying that cohesion in video discourse should extend beyond linguistic ties to include visual and temporal connections such as sequencing, framing, and audiovisual alignment.

J. Liu et al. (2025) support this claim empirically, showing that spoken communication in video formats relies heavily on lexical repetition, personal reference, and discourse markers to sustain coherence and audience engagement, though their analysis rarely operationalizes the visual dimension

that Jewitt et al. (2016) foreground, focusing instead on verbal cohesion alone. Mayer (2020) and Noetel et al. (2021) approach the same video-based tasks from yet another angle, emphasizing that they promote authentic language use by requiring learners to organize ideas and structure narratives, but without offering detailed linguistic analysis of how cohesion is achieved. More directly relevant, Ngo (2026) develops a systemic functional semiotic account of film sound, demonstrating that music and sound effects can themselves function cohesively within multimodal texts. This work offers a systematic, device-level framework for describing cross-modal cohesion, though it is developed for edited cinematic texts produced under different technical conditions than the spontaneous, unedited spoken videos examined here. Taken together, these four strands of research point in a consistent theoretical direction, that cohesion in video discourse is multimodal, without yet providing the empirical, device-level analysis of student-generated video needed to substantiate it. This study addresses that gap directly by coding both verbal and visual cohesive devices within the same dataset.

ESP and Digital Communication Competence

In the context of ESP, language learning is closely aligned with the communicative demands of specific professional fields (Hutchinson & Waters, 1991). ESP research emphasizes that learners must develop not only grammatical accuracy but also discourse competence, including the ability to construct coherent and contextually appropriate communication. For communication studies students, this involves producing persuasive and structured messages in digital formats, such as video presentations, interviews, and social media content.

Pamuji et al. (2025) support this claim at a general level, showing that digital project-based learning enhances students' communicative competence by integrating language use with real-world tasks. Setiawan et al. (2025) extend this argument further, contending that ESP instruction in digital contexts should center on authentic communication practices in which learners actively construct meaning and engage audiences. Yet both strands of this literature describe discourse competence largely in outcome terms, as something students demonstrate or instruction should target, without specifying the linguistic mechanisms through which that competence is enacted. Cohesion, as a concrete and analyzable set of devices, offers precisely such a mechanism, providing an empirical bridge between ESP's pedagogical concerns and the discourse-analytic frameworks discussed earlier in this review.

METHOD

Research Design

This study employs a qualitative research design grounded in discourse analysis to examine the use of cohesion in student-generated video discourse. A qualitative approach is appropriate as it enables an in-depth exploration of linguistic patterns and meaning-making processes in naturally occurring communication rather than relying on numerical generalization (Creswell & Poth, 2025). The study specifically adopts a discourse-analytic perspective to

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

investigate how cohesive devices are used to organize ideas, construct meaning, and maintain coherence in spoken English.

The analytical framework is informed by Halliday and Hasan's theory of cohesion, complemented by a Systemic Functional Linguistics (SFL) perspective which conceptualizes language as a social semiotic system (Halliday & Matthiessen, 2014). This combined framework allows cohesion to be examined not only as a structural feature but also as a functional resource for communication. In addition, the study is conceptually aligned with discourse-analytic traditions that view language as social practice embedded in context (Fairclough, 2013; Gee, 2014), enabling the analysis to capture how students construct meaning within a digital communication environment.

Research Setting

The study was conducted in a higher education context, specifically within an English for Communication course designed for Communication Studies students at Universitas Ahmad Dahlan, Indonesia. The course adopts an ESP approach, emphasizing the development of communicative competence in digital and professional contexts.

The research setting involves a digital project-based learning environment in which students are required to produce English self-introduction videos as part of their coursework. These videos function as authentic communicative artifacts that integrate spoken language, visual presentation, and media production. Such a setting reflects real-world communication practices, thereby enhancing the ecological validity of the study (Braun & Clarke, 2022). It also aligns with contemporary perspectives in applied linguistics, which emphasize the importance of analyzing language use in multimodal and digitally mediated contexts.

Data and Participants

The data for this study consist of 13 English self-introduction videos produced by undergraduate communication studies students. Rather than focusing on participants as individuals, this study treats each video as a unit of discourse, consistent with discourse analysis approaches that prioritize texts or spoken outputs as primary data (McEnery & Hardie, 2012).

Each video has a duration of approximately 2–3 minutes and represents an authentic performance of spoken English in a structured communicative task. The selection of data followed purposive sampling criteria to ensure comparability and relevance: (1) the video must be delivered in English, (2) the speaker must be enrolled in the English for Communication course, and (3) the video must present a self-introduction in a coherent format. These criteria ensure that the dataset reflects a consistent genre of discourse suitable for cohesion analysis.

Data collection and analysis proceeded iteratively, and no new cohesive device types or functional patterns emerged after the thirteenth video was analyzed, indicating that data saturation had been reached. Given the relatively homogeneous genre and task constraints of the self-introduction videos (a fixed assignment prompt, comparable duration, and a shared communicative purpose), this sample size was judged sufficient to capture the range of cohesive

strategies typically employed by students in this task. A larger or more varied corpus may nonetheless reveal additional patterns, a limitation acknowledged in the conclusion.

Data Collection

The data were collected from course assignments submitted by students during the semester. Prior to analysis, all videos were transcribed verbatim to capture the linguistic features of spoken discourse, including pauses, repetitions, and discourse markers. Transcription is a crucial step in discourse analysis as it allows for systematic examination of language use at both micro and macro levels (Gee, 2014).

In addition to the verbal transcripts, visual elements such as gesture, on-screen framing, and sequencing were noted informally during repeated viewing of the videos. These observations were not coded systematically or analyzed as an independent data layer. Rather, they were recorded when they appeared to reinforce or accompany a verbal cohesive device already identified in the transcript, for example, a demonstrative reference co-occurring with a pointing gesture. This study therefore does not claim to provide a systematic multimodal coding of the videos; visual observations serve an illustrative and interpretive function, supplementing the primary linguistic analysis of cohesion rather than constituting a parallel, independently coded dataset.

Data Analysis

The data analysis was conducted using a qualitative coding procedure based on the cohesion framework of Halliday and Hasan. The analysis proceeded in two main stages.

First, all instances of cohesive devices were identified and categorized into grammatical cohesion (reference, substitution, ellipsis, conjunction) and lexical cohesion (repetition, synonymy, antonymy, and collocation). Second, each instance was analyzed functionally to examine how it contributes to discourse coherence, clarity, and communicative effectiveness.

The analysis followed an iterative process involving repeated reading of transcripts, coding refinement, and pattern identification across the dataset. This approach allows for the development of thematic insights rather than merely descriptive counts, aligning with qualitative data analysis principles (Miles et al., 2014). Through this process, the study aims to reveal not only what cohesive devices are used but also how they function in shaping spoken discourse in digital contexts.

It should be noted that this analytic procedure canters on linguistic cohesion as coded within Halliday and Hasan's framework. Visual and other semiotic elements were not subjected to an independent, systematic coding scheme, such as a formal multimodal transcription protocol. Instead, they were observed impressionistically alongside the linguistic coding and are reported only in instances where they visibly co-occurred with, and appeared to support, an already-identified verbal cohesive device. Accordingly, references to multimodal cohesion in this study denote instances where a verbal cohesive device co-occurs with a supporting visual cue, rather than the product of a fully

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

systematic cross-modal analysis. This narrower scope is treated as a deliberate boundary of the present study, and more systematic multimodal coding is recommended for future research.

Trustworthiness

To ensure the credibility and trustworthiness of the findings, several strategies were employed. First, data triangulation was achieved by analyzing multiple video samples, allowing for cross-case comparison. Second, informal peer debriefing was conducted: a colleague with expertise in discourse analysis reviewed a subset of the coded transcripts and discussed ambiguous or borderline classifications with the first author, and disagreements were resolved through discussion and refinement of the coding categories. Because coding was carried out by a single researcher and peer debriefing was informal rather than based on independent double-coding, no quantitative inter-coder agreement statistic, such as Cohen's kappa, was calculated. This is acknowledged as a limitation of the study's analytical procedure. Third, transparency was maintained through detailed documentation of analytical procedures and representative data excerpts.

In qualitative discourse analysis, trustworthiness is established through depth of analysis, consistency of interpretation, and alignment with theoretical frameworks rather than statistical generalization (Creswell & Poth, 2025). Therefore, this study emphasizes analytical rigor and methodological transparency to enhance the reliability of its findings.

RESULTS AND DISCUSSION

Results

The analysis of 13 English self-introduction videos produced by communication studies students reveals a coherent and systematically organized set of cohesive strategies. Rather than reflecting incomplete grammatical control, the distribution of cohesive devices demonstrates that students actively prioritize communicative effectiveness within the constraints of real-time spoken production. A total of 90 instances of cohesive devices were identified across the dataset, distributed across five dominant categories: personal reference (25 instances, 27.8%), conjunction (23 instances, 25.6%), lexical repetition (22 instances, 24.4%), demonstrative reference (14 instances, 15.6%), and collocation (6 instances, 6.7%). The following table presents the complete distribution of cohesive devices across each video.

Table 1. Distribution of Cohesive Devices Across the 13 Videos

Video	Personal Ref.	Demonstrative Ref.	Conjunction	Lexical Repetition	Collocation	Total
Video 1	3	1	2	2	1	9
Video 2	3	2	2	2	1	10
Video 3	2	1	3	3	1	10
Video 4	2	1	2	1	0	6
Video 5	2	1	1	2	0	6
Video 6	2	1	1	1	1	6
Video 7	1	0	1	1	0	3
Video 8	1	3	1	2	0	7
Video 9	3	2	3	3	1	12
Video 10	2	0	2	1	0	5

Video 11	1	1	1	2	0	5
Video 12	2	0	2	1	1	6
Video 13	1	1	2	1	0	5
Total	25	14	23	22	6	90
Percent age	27.8%	15.6%	25.6%	24.4%	6.7%	100%

As shown in Table 1, personal reference is the most frequently occurring cohesive device, appearing 25 times across all 13 videos. Conjunction follows closely with 23 instances, while lexical repetition accounts for 22 occurrences, indicating that grammatical and lexical devices are deployed with approximately equal frequency. Demonstrative reference appears 14 times, predominantly in videos that involve physical space or object-oriented narration (notably Videos 2, 8, and 9). Collocation, while the least frequent at 6 instances, is concentrated in videos with a defined thematic field, particularly those related to design (Video 1), cosmetics (Video 6), and literary discussion (Video 9). These figures establish the quantitative basis from which the following functional analysis proceeds.

Table 2. Illustrative Examples of Dominant Cohesive Devices

Category	Device	Utterance	Function
Grammatical	Personal Reference (Anaphoric)	"At first, I was just another student... my journey started..."	Anchors speaker identity; maintains referential continuity across the video.
Grammatical	Personal Reference (Inclusive 'we')	"We learn to be responsible when we meet deadlines."	Shifts from monologue to dialogic positioning; aligns speaker with audience.
Grammatical	Demonstrative Reference ('This')	"This is my story." / "This is the place."	Bridges verbal and visual modes; creates multimodal cohesion.
Grammatical	Conjunction (Causal)	"I went to the mountain because I needed silence."	Encodes causal relation; sustains logical coherence.
Grammatical	Conjunction (Adversative)	"It's not about following trends, but about feeling good."	Signals contrast; reinforces the speaker's stance.
Grammatical	Conjunction (Conditional)	"So, if life ever gets loud, try going somewhere quiet."	Connects premise to recommendation; guides audience reasoning.
Grammatical	Ellipsis (Verbal)	"I smile, tear up, even pause to breathe..."	Omits repeated operators; increases fluency and rhythmic flow.
Grammatical	Comparative Reference	"Character matters more than grades."	Introduces evaluative meaning; moves beyond description.
Lexical	Repetition	"Fashion... lifestyle... fashion aesthetics... fashion..."	Sustains thematic chain; guides audience interpretation.
Lexical	Repetition	"Mountain... mountain never demands... mountain never asks..."	Foregrounds central theme; reinforces rhetorical effect.
Lexical	Synonym/Near-Synonym	"escape ~ run away"; "feel like myself ~ stay true to who I am"	Maintains cohesion through semantic variation; avoids direct repetition.
Lexical	Superordinate	"Before the Coffee Gets Cold" → "novel" → "books"	Links specific to general; taxonomic cohesion across discourse.

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

Lexical	Collocation	<i>"design, spacing, contrast, visual, Canva"</i>	Domain-specific cluster; constructs professional identity through lexical field.
Lexical	Collocation	<i>"moisturizer, foundation, blush, lipstick, mascara"</i>	Cosmetic field cohesion; authenticates speaker expertise.

Table 2 presents representative utterances illustrating the functional role of each cohesive category. The examples are drawn directly from the video transcripts and demonstrate how each device contributes to discourse organization, thematic continuity, and audience engagement. The following subsections provide a detailed functional analysis of each pattern.

Personal Reference as the most Frequent Cohesive Device

Personal reference emerges as the most frequently occurring cohesive device across the dataset, appearing in every video and accounting for the largest share of coded instances (27.8%, Table 1). First-person pronouns (*I, me, my*) function not simply as grammatical referential markers but as anchors of textual continuity and identity construction. Utterances such as *"At first, I was just another student"* (Video 1) and *"My purpose is to bring happiness to others through my music"* (Video 10) demonstrate how learners maintain referential tracking while simultaneously constructing a coherent narrative voice. In both cases, the pronoun *I* creates an anaphoric chain that links successive propositions and positions the speaker as the central agent of the discourse. This high frequency suggests that personal reference plays an important role in sustaining referential continuity; however, the present data, which report distributional frequency rather than a direct test of organizing function, are not sufficient to establish personal reference as the primary mechanism of discourse organization overall. Establishing that claim would require additional evidence, for example, examining whether removing or substituting these devices disrupts perceived coherence, or systematically comparing their organizing contribution against that of lexical and conjunctive devices. The discussion below therefore focuses on how personal reference is used, rather than asserting its centrality relative to other cohesive resources.

Beyond individual identity, the data also reveal a strategic use of inclusive and second-person reference as an interpersonal cohesive resource. In expressions such as *"We learn to be responsible when we meet deadlines"* (Video 4) and *"What do you remember?"* (Video 7), personal reference is mobilized to align the speaker with the audience, transforming the discourse from monologic narration into dialogic positioning. This finding extends Halliday and Hasan (2014) conceptualization of personal reference as a textual device, demonstrating that in spoken digital discourse, reference simultaneously performs textual and interpersonal metafunctions.

Lexical Cohesion as a Strategy for Thematic Control

Lexical repetition functions as the central strategy for maintaining thematic control across individual videos. The recurring use of key terms forms cohesive chains that stabilize meaning across extended stretches of discourse, enabling audiences to follow the speaker's thematic trajectory without explicit structural signaling. In Video 2, the lexical item *fashion* recurs no fewer than six times across three discourse segments, functioning as a lexico-thematic anchor that

prevents topic drift. Similarly, in Video 3, *mountain* is repeated systematically across successive clauses, "*The mountain never demands strength. It never asks you to rush. It simply waits*", forming a syntactically parallel lexical chain that reinforces the central rhetorical argument.

Beyond simple repetition, more sophisticated forms of lexical cohesion are evident in several videos. In Video 9, the discourse transitions progressively from the specific term *Before the Coffee Gets Cold* to the near-synonymous *novel* and subsequently to the superordinate *books*, creating a taxonomic cohesive chain that structures the speaker's comparative framework. Domain-specific collocational clusters are observable in Videos 1 (*design, spacing, contrast, visual, Canva*), Video 6 (*moisturizer, foundation, blush, lipstick, mascara*), and Video 9 (*cafe, rules, seat, coffee gets cold*), where co-occurring lexical items from a shared semantic field collectively reinforce the topic and authenticate the speaker's subject-matter knowledge.

Demonstrative Reference and Multimodal Cohesion

A particularly significant pattern in the data is the use of demonstrative reference (*this, that, these, there*) as a bridge between verbal and visual modes of communication. Unlike the anaphoric pronoun chains characteristic of personal reference, demonstrative forms in the video corpus frequently operate exophorically, presupposing a shared visual frame rather than a linguistic antecedent within the text. In Video 8, the speaker's repeated use of "*This is the place*" and "*This is basically the pilot in the place*" acquires full referential meaning only in conjunction with the camera-captured environment. Detached from the visual channel, the demonstrative *this* would remain semantically incomplete.

Similarly, in Video 1, "*That's when everything started to change*" functions both linguistically, referring anaphorically to the narrated event of winning a competition, and visually, as the shift in the speaker's tone and posture appears to reinforce the demonstrative's temporal deixis. These examples are illustrative rather than the product of a systematic visual coding scheme (see Method), and they are best read as suggestive instances rather than dataset-wide evidence. Even on this limited basis, however, they indicate that cohesion in video discourse is not necessarily confined to language alone and may, in at least some cases, be co-constructed through the interaction of speech and image, a pattern that warrants more systematic investigation in future research using a dedicated multimodal coding protocol.

Conjunctions and the Minimalist Logic of Coherence

Conjunctions constitute the second most frequently used cohesive category in the dataset, with 23 instances distributed across all 13 videos. However, a notable characteristic of their use is the concentration of a small, high-frequency repertoire rather than a wide range of connective types. The adversative conjunction *but*, the causal conjunctions *because* and *so*, and conditional constructions with *if* collectively account for the majority of conjunctive instances. This repertoire, while limited in syntactic variety, proves functionally sufficient for organizing discourse across additive, causal, adversative, and conditional domains.

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

Causal relations are particularly clearly expressed: utterances such as *"I went to the mountain because I needed silence"* (Video 3) and *"I was very shy, so I started posting videos"* (Video 10) encode reason-result relations with minimal syntactic complexity. Adversative conjunctions are deployed to signal rhetorical contrast and speaker stance: *"It's not about following trends, but about feeling good and confident"* (Video 2), and *"not to escape, but to be still"* (Video 3).

Substitution, Ellipsis, and Markers of Emerging Discourse Efficiency

Although less frequent than reference and lexical devices, instances of substitution and ellipsis provide evidence of developing discourse efficiency in the student corpus. Nominal substitution is observed in Video 9, where *"before this"* substitutes for the full noun phrase *this Japanese novel*, and in Video 7, where *"the book"* functions as a nominal substitute for the previously mentioned *notes*. Verbal ellipsis is evident in Video 9 (*"I smile, tear up, even pause to breathe"*), where the subject and auxiliary are omitted in successive predicates, and in Video 3 (*"slow down, but don't stop"*), where modal ellipsis omits the subject and finite operator. These constructions indicate a developing sensitivity to informational redundancy and an emerging capacity for economical expression, which are characteristic features of more advanced spoken discourse.

Comparative Structures and Evaluative Meaning-Making

Comparative cohesive devices, while not the most frequent category, represent an analytically significant dimension of the data because they signal a qualitative shift in discourse function. Utterances such as *"Character matters more than grades"* (Video 4), *"Makeup is more than a routine"* (Video 6), and *"The real reward is not the summit, it's the journey"* (Video 3) demonstrate learners' capacity to move beyond descriptive narration into evaluative and argumentative discourse. In these instances, comparison functions not merely as a cohesive tie but as a rhetorical device for constructing stance, encoding value judgments, and guiding audience interpretation.

The more complex comparative construction in Video 3, *"The real reward is not the summit, it's the journey"*, illustrates an ability to reframe meaning through contrastive structure, a discourse strategy associated with persuasive and reflective registers. This indicates that a subset of students is beginning to exploit cohesion as a resource for constructing evaluative discourse, suggesting that communicative competence in this domain extends beyond structural accuracy to encompass rhetorical organization.

Discussion

The findings of this study offer several significant theoretical and pedagogical insights into the nature of cohesion in student-generated multimodal discourse. Drawing on Halliday and Hasan (2014) foundational framework, the analysis confirms that both grammatical and lexical cohesive devices are systematically deployed in spoken digital communication; however, the results also reveal patterns that cannot be adequately explained by traditional text-based models of cohesion alone. The following discussion addresses the broader implications of these findings across three interrelated dimensions: the reconceptualization of cohesion as a functional competence, the multimodal extension of cohesive meaning-making, and the pedagogical consequences for ESP instruction.

A contribution of this study is to offer a counterpoint to a deficit-oriented reading of limited cohesive range in second language learner discourse, an interpretation illustrated in prior work associating a narrow repertoire of cohesive devices with lower discourse quality (Crossley et al., 2016). Rather than claiming that such a deficit orientation has broadly characterized the field as a whole, this study addresses that specific interpretive tendency directly: the strategic concentration of a small, high-yield repertoire of cohesive devices in the present dataset reflects a principled adaptation to the constraints of spontaneous spoken production rather than a gap in linguistic knowledge. The dominance of personal reference (27.8%) and its functional deployment for both identity construction and interpersonal alignment, as evidenced by the strategic alternation between *I*, *we*, and *you*, indicates that learners are not merely using cohesion to connect clauses but are employing it as a resource for negotiating social relationships and positioning themselves within the communicative event. This is consistent with Hyland's (2019) argument that discourse is fundamentally interpersonal; the present findings illustrate this perspective in the context of student video discourse, showing personal reference operating across textual and interpersonal metafunctions simultaneously in this performance-based digital communication context.

The reinterpretation of lexical repetition as thematic strategy rather than vocabulary limitation constitutes a second significant contribution. While repetition is frequently cited as a characteristic of lower-proficiency learner output, its systematic deployment in the present corpus, particularly the formation of extended lexical chains around core thematic items such as *fashion*, *mountain*, and *book*, reveals a coherence-oriented communicative strategy. From a processing perspective, repeated lexical items reduce cognitive load for both the producer and the receiver, enabling the speaker to maintain thematic focus under real-time production pressure while simultaneously guiding audience interpretation. This aligns with and extends the findings of Crossley et al. (2016), who identified lexical cohesion as a predictor of discourse quality, by demonstrating that repetition specifically serves a thematic anchoring function in spoken multimodal contexts. Consequently, any pedagogical framework that categorizes repetition as an error of avoidance rather than a functional strategy would misrepresent the communicative sophistication underlying the students' discourse choices.

A further contribution of this study is a preliminary observation suggesting that cohesion in video discourse may not be confined to language alone. Traditional cohesion theory, as developed by Halliday and Hasan, is grounded in written and spoken linguistic data and does not directly address the integration of visual and auditory semiotic resources. In the present dataset, a small number of demonstrative references such as *this* and *that* appeared to presuppose a shared visual frame rather than a linguistic antecedent, suggesting that in these specific instances the cohesive tie may have been distributed across verbal and visual channels rather than residing in language alone. Because these observations were noted informally rather than through a systematic multimodal coding scheme (see Method), they should be read as an illustrative possibility rather than a confirmed finding, and they do not, on their

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

own, establish that cohesion in video discourse is inherently multimodal. This tentative observation nonetheless points to a direction for future research: a systematic, dedicated coding of visual and verbal cohesive resources would be needed to determine whether this pattern reflects a broader, generalizable feature of video-mediated cohesion. Emerging frameworks such as Ngo (2026) systemic functional semiotic account of cohesion in film sound offer one promising methodological direction for such future coding, though extending that framework, developed for edited cinematic texts, to unedited, learner-produced video would require substantial adaptation.

The findings regarding conjunction use further nuance existing accounts of coherence in learner discourse. The dominance of a small number of high-frequency connectors, predominantly *but*, *because*, and *so*, across all videos suggests that students have internalized what can be described as a minimalist logic of coherence, whereby communicative clarity is achieved through efficient rather than elaborate connective means. This is consistent with McNamara et al.'s (2014) computational analysis, which found that effective discourse does not necessarily rely on a wide range of surface cohesive markers; rather, coherence depends on the semantic integration of propositional content. The present data extend this finding into authentic spoken ESP contexts, demonstrating that even a restricted conjunctive repertoire can sustain discourse coherence when deployed in alignment with clear thematic structure and communicative intent.

The use of evaluative discourse through comparative structures represents an analytically noteworthy feature of the students' communicative repertoire, as observed in this dataset. Constructions such as "*Character matters more than grades*" and "*The real reward is not the summit, it's the journey*" indicate that some learners deploy cohesion not only as a linking mechanism but also as a resource for constructing rhetorical stance and argumentation. This finding is theoretically relevant because it suggests that cohesive competence, as observed here, involves not only the use of grammatical forms but also the capacity to exploit cohesive devices for evaluative and persuasive purposes, a dimension that has received limited attention in learner discourse research. As this study is based on a single set of videos collected at one point in time, these patterns should be understood as features observed in the current dataset rather than evidence of developmental change; longitudinal research would be needed to examine whether, or how, such evaluative use of cohesion develops over time.

From a pedagogical perspective, these findings have direct implications for ESP instruction within communication studies departments. The evidence that students effectively deploy cohesion as a functional communicative strategy rather than merely a grammatical requirement suggests that instruction should prioritize discourse-based approaches that develop learners' awareness of how cohesive devices contribute to meaning-making in authentic communicative contexts. Specifically, instructional activities should integrate spoken production tasks that require learners to manage thematic continuity, construct audience alignment, and coordinate verbal and visual resources within a single communicative performance. Video-based project assignments of the kind examined in this study are well-suited to this purpose, as they provide a

meaningful communicative context in which learners must deploy cohesion strategically rather than mechanically. Furthermore, the multimodal dimension of cohesion identified in this study suggests that ESP instruction should extend beyond linguistic cohesion to encompass the analysis and production of multimodal discourse, equipping learners with the resources needed to communicate effectively in the digital professional environments they will encounter as communication studies graduates.

Taken together, these findings support a reconceptualization of cohesion in applied linguistics as a dynamic, adaptive, and multimodal communicative resource rather than a fixed structural property of text. The study proposes that cohesion in student video discourse is organized around three interrelated functional dimensions: identity-oriented cohesion, realized through personal reference and inclusive pronouns; theme-oriented cohesion, achieved through lexical repetition, superordinates, and collocational chains; and logic-oriented cohesion, structured through conjunctions and comparative constructions. These dimensions do not operate in isolation but interact dynamically within the multimodal communicative event, suggesting that future frameworks for analyzing cohesion in digital discourse must account for both linguistic and semiotic dimensions of meaning-making.

CONCLUSION

This study examined how cohesion is realized in English project videos produced by communication studies students in an ESP context. The findings show that personal reference, lexical repetition, and conjunction were the dominant cohesive devices in this dataset, used strategically to organize ideas and sustain coherence under real-time spoken conditions rather than reflecting grammatical complexity. In a limited number of instances, cohesion also appeared to draw on visual as well as verbal resources, though this observation was based on informal rather than systematic multimodal analysis and should be treated as preliminary.

These findings suggest that ESP instruction may benefit from complementing a focus on grammatical accuracy with discourse-based approaches that address how cohesion functions in authentic, digital communicative tasks. Because this study draws on a small qualitative dataset from a single institutional context, coded by a single researcher, these findings are best regarded as illustrative rather than generalizable. Larger, multi-institutional datasets and systematic multimodal coding would be needed to test whether the patterns identified here hold more broadly.

AUTHOR STATEMENTS

Prayudha, as the corresponding author and principal investigator, led the research conceptualization, design, data collection, analysis, manuscript writing, and supervision. **Rahmi Munfangati** contributed to literature review development, methodological refinement, and manuscript review, while **Fitri Heyza Galuh Kirana** supported data transcription, coding, and research administration. AI tools were used to assist with language refinement, structural editing, and manuscript preparation, with all outputs critically

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

reviewed and validated by the authors, who take full responsibility for the final scholarly work.

ACKNOWLEDGEMENTS

The data for this study were collected through research funded by the Directorate of Research and Community Service (DPPM 2025), under the Ministry of Higher Education, Science, and Technology of Indonesia. The author gratefully acknowledges the financial support and institutional assistance provided for this research.

REFERENCES

- Alfaridhi, D. F., Prayudha, P., Ali, R. M., & Masduki, A. (2026). Students' perceptions of an ai-based English e-module for digital content creation: An explanatory sequential mixed-methods study. *Journal of Languages and Language Teaching*, 14(1), 299–311. <https://doi.org/10.33394/jollt.v14i1.17752>
- Barkaoui, K. (2026). Quality in second language classroom assessment: Controversies, challenges, and the need for research on teacher conceptions and practices. *Language Assessment Quarterly*, 23(1), 1–23. <https://doi.org/10.1080/15434303.2026.2620813>
- Barrot, J. S. (2022). Social media as a language learning environment: A systematic review of the literature (2008-2019). *Computer Assisted Language Learning*, 35(9), 2534–2562. <https://doi.org/10.1080/09588221.2021.1883673>
- Braun, Virginia., & Clarke, Victoria. (2022). *Thematic analysis: A practical guide*. SAGE.
- Bruce, T. (2018). New technologies, continuing ideologies: Online reader comments as a support for media perspectives of minority religions. *Discourse, Context & Media*, 24, 53–75. <https://doi.org/10.1016/j.dcm.2017.10.001>
- Creswell, J. W., & Poth, C. N. (2025). *Qualitative inquiry & research design: Choosing among five approaches*. Sage.
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The development and use of cohesive devices in L2 writing and their relations to judgments of essay quality. *Journal of Second Language Writing*, 32, 1–16. <https://doi.org/10.1016/j.jslw.2016.01.003>
- D'Angelo, L., & Maci, S. (2021). Metadiscourse in digital communication: A short introduction. In *Metadiscourse in Digital Communication* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-030-85814-8_1
- Dijk, T. A. van. (2014). *Discourse and knowledge*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107775404>
- Fairclough, Norman. (2013). *Critical discourse analysis: The critical study of language*. Routledge.
- Fairclough, Norman. (2015). *Language and power* (rd). Routledge, Taylor & Francis Group.
- Flowerdew, J. (2012). *Discourse in English language education*. Routledge. <https://doi.org/10.4324/9780203080870>

- Gee, J. P. (2014). *An Introduction to discourse analysis*. Routledge.
<https://doi.org/10.4324/9781315819679>
- Hafner, C. A., & Ho, W. Y. J. (2020). Assessing digital multimodal composing in second language writing: Towards a process-based model. *Journal of Second Language Writing*, 47, 100710. <https://doi.org/10.1016/j.jslw.2020.100710>
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar*. Routledge.
- Halliday, M. A. K., & Hasan, R. (2014). *Cohesion in English*. Routledge.
<https://doi.org/10.4324/9781315836010>
- Hutchinson, & Waters. (1991). *English for specific purposes: A learning-centred approach*. Cambridge University Press.
- Hyland, Ken. (2019). *Metadiscourse: Exploring interaction in writing*. Bloomsbury Academic.
- Jewitt, C., Bezemer, J., & O'Halloran, K. (2016). *Introducing multimodality*. Routledge.
<https://doi.org/10.4324/9781315638027>
- Jiang, L., & Luk, J. (2016). Multimodal composing as a learning activity in English classrooms: Inquiring into the sources of its motivational capacity. *System*, 59, 1–11. <https://doi.org/10.1016/j.system.2016.04.001>
- Jones, R. H., Chik, A., & Hafner, C. A. (2015). *Discourse and digital practices*. Routledge.
<https://doi.org/10.4324/9781315726465>
- Lindenberg, D. (2023). Modes and intersemiotic cohesion in student presentations performed online: An SF-informed multimodal discourse analysis. *English for Specific Purposes*, 69, 67–79. <https://doi.org/10.1016/j.esp.2022.10.002>
- Liu, H., Liu, L., & Li, H. (2024). Multimodal discourse studies in the international academic community (1997–2023): A bibliometric analysis. *Sage Open*, 14(4).
<https://doi.org/10.1177/21582440241305454>
- Liu, J., Shi, Z., & Li, W. (2025). Understanding genre differences in cohesive use of EFL Learners: Insights from systemic functional linguistics. *System*, 132, 103692.
<https://doi.org/10.1016/j.system.2025.103692>
- Lorenzo-Dus, N., Izura, C., & Pérez-Tattam, R. (2016). Understanding grooming discourse in computer-mediated environments. *Discourse, Context & Media*, 12, 40–50. <https://doi.org/10.1016/j.dcm.2016.02.004>
- Mayer, R. (2020). *Multimedia learning*. Cambridge University Press.
<https://doi.org/10.1017/9781316941355>
- McEnery, Tony., & Hardie, Andrew. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
<https://doi.org/10.1017/CBO9780511894664>
- Miles, M. B., Huberman, A. M., & Saldaña, Johnny. (2014). *Qualitative data analysis: A methods sourcebook*. SAGE Publications, Inc.
- Ngo, T. (2026). Language of film sounds: Developing a metalanguage for talking about film sounds as literary devices in the language arts classes. *Journal of World Languages*, 12(1), 185–209. <https://doi.org/10.1515/jwl-2025-0042>

How to Cite (APA Style):

Prayudha, P., Munfangati, R., & Kirana, F.H.G. (2026). Cohesion as functional resource in multimodal ESP discourse: A study of communication studies students' English project videos. *EduLite: Journal of English Education, Literature, and Culture*, 11 (2), 559-576. <http://dx.doi.org/10.30659/e.11.2.559-576>

Noetel, M., Griffith, S., Delaney, O., Sanders, T., Parker, P., del Pozo Cruz, B., & Lonsdale, C. (2021). Video improves learning in higher education: A systematic review. *Review of Educational Research*, 91(2), 204–236. <https://doi.org/10.3102/0034654321990713>

Pamuji, Kristanto, W., Efendi, T. A., & Khotimah, N. (2025). Integrating project-based learning and video exposé to enhance English language teaching in higher education: A case study of student empowerment. *Cogent Education*, 12(1). <https://doi.org/10.1080/2331186X.2025.2578371>

Prayudha, P. (2016). The cohesion and coherence of the editorials in The Jakarta Post. *Ahmad Dahlan Journal of English Studies*, 3(2), 30. <https://doi.org/10.26555/adjes.v3i2.4987>

Prayudha, P., Ali, R. M., Masduki, A., & Alfaridhi, D. F. (2026). AI-based English module needs analysis for Indonesian content creator. *English Language Teaching Educational Journal*, 9(1), 132–142. <https://doi.org/10.12928/eltej.v9i1.14893>

Setiawan, A., Hartono, R., Suwandi, & Fitriati, S. W. (2025). English for specific purposes (ESP) in the digital era: Bibliometric mapping with critical reflection on the lack of performance-based learning studies. *Multidisciplinary Reviews*, 9(2), 2026050. <https://doi.org/10.31893/multirev.2026050>

Wigham, C. R., & Satar, M. (2021). Multimodal (inter)action analysis of task instructions in language teaching via videoconferencing: A case study. *ReCALL*, 33(3), 195–213. <https://doi.org/10.1017/S0958344021000070>

Yu, H., & Lowie, W. (2020). Dynamic paths of complexity and accuracy in second language speech: A longitudinal case study of Chinese learners. *Applied Linguistics*, 41(6), 855–877. <https://doi.org/10.1093/applin/amz040>

Zhou, T., Cao, S., Zhou, S., Zhang, Y., & He, A. (2023). Chinese intermediate English learners outdid ChatGPT in deep cohesion: Evidence from English narrative writing. *System*, 118, 103141. <https://doi.org/10.1016/j.system.2023.103141>

Conflict of Interest Statement: The authors declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright©2026. **Prayudha, Munfangati, and Kirana.** This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution, or reproduction is permitted that does not comply with these terms.