



STUDENTS' PERCEPTIONS OF L1 VS. L2 CAPTIONS IN VIDEO-BASED LISTENING AT A VIETNAMESE UNIVERSITY

Nguyen Thi Bach Lienⁱ

Gia Dinh University,
Vietnam

Abstract:

The present study explored the perceptions of Vietnamese university students about the use of first (L1) and second language (L2) captions in video input in order to establish how these captions can mediate their cognitive processing of and motivation for listening. The study utilized a sequential explanatory mixed-methods design, with survey data from 130 English majors supplemented by interviews with six students. Results showed that students attributed different functional uses to each form of captionation: L1 captions were preferred where emotional reassurance and comprehension automatically play the roles of emotion and cognition, respectively, while L2 captions led to recognition of language gains and self-validation. Rather than preferring one of the media, the respondents used them interchangeably as a function of goals sought. These findings illustrate that caption use is a metacognitive function rather than a technical one, and indicate that flexible integration of captions should be the goal of listening pedagogy.

Keywords: perceptions, video-based, captions, listening skills, Vietnamese context

1. Introduction

Listening has long been known as a primary skill in learning a second language, accounting for nearly half of the communicative activity on a day-to-day basis (Rost, 2011). The rapid growth of multimedia and streaming services has seen video as one of the most common means in which students receive authentic input in spoken English (Wang & Chen, 2020). Unlike traditional audio forms of input, video, with its added advantage of visual imagery and subtitles, provides a multimodal environment from which comprehension may be enhanced in terms of ease of understanding and motivation. As a result, the teaching potential of captioned video listening has emerged as a welcome pedagogical tool in the field of English language education throughout the world.

ⁱ Correspondence: email leslienguyen0310@gmail.com

One important aspect of the planning of captioned listening materials is whether the captions should be in the first language (L1) or the second language (L2) of the learners. On the one hand, captions in L1 may encourage faster understanding and less cognitive interference, whereas captions in L2 would seem to provide greater exposure to linguistic features as well as vocabulary acquisition (Vanderplank, 2016). In Vietnamese tertiary EFL contexts, both kinds of captioning are commonly found on self-study systems, in language centers, and in listening tasks in the classroom, but the choice of L1 or L2 appears to be more a decision made from teacher preference than from the needs of the learners.

Various studies have produced mixed results with regard to the effectiveness and learner preference of L1 or L2 captions. It appears that the L1 might work better in providing confidence and initial understanding to less able learners (Montero Perez *et al.*, 2014), whereas the L2 terms might produce greater processing ability and long-term retention of the language (Winke *et al.*, 2013). Again, it is suggested that captions lead to greater attention, less listening anxiety, and longer sustained independent learning (Li, 2022; Lee *et al.*, 2021). All these suggestions point to captioning as a choice that is not only technical in nature but also has pedagogical implications psychologically.

Most previous research, however, has been conducted in Western or high-exposure ESL settings, while research in Asian EFL contexts, particularly Vietnam, is rare. More importantly, previous research has tended to focus on performance outcomes rather than students' subjective perceptions, preferences, and self-reported experiences. Little empirical evidence exists on how Vietnamese university students perceive the cognitive effort, usefulness, and motivational implications of either L1 or L2 captions when engaged in video-based listening.

Perceptions of students are important to understand if informed pedagogical decisions are to result in appropriate technology-assisted listening instruction. The outcomes of the current study will benefit not only language teachers who are trying to optimize the use of captions in the classroom, but also material designers, platform developers, and students themselves, who can make more informed and strategic choices in their independent listening practice.

The present study investigates students' subjective experiences of two commonly used caption types in video-based listening. Specifically, it will be guided by the following research questions:

- 1) What are Vietnamese university students' perceptions of L1 captions compared with L2 captions in terms of perceived usefulness, cognitive effort, and motivational impact during video-based listening?
- 2) How do students explain their preferences for either L1 or L2 captions when engaging with video-based listening materials?

2. Literature review

2.1. The role of captions in L2 listening comprehension

Listening is often identified as the most difficult of the four language skills for EFL learners, due in part to the ephemeral nature of spoken input and also because of the limited means to manipulate instant replay in real-time communication (Field 2008). Unlike reading, listening requires quick processing on many levels: phonological decoding, access to vocabulary, syntactic parsing, and integration of meaning, all of which represent a substantial cognitive load for learners (Vandergrift & Goh, 2012). Thus, instructional interventions that can scaffold the process of understanding are necessary for successful listening development.

Captions, described as written texts on screen that are synchronized with audio input, are being acknowledged increasingly as an important form of multimodal support in technology-facilitated listening instruction. Underpinning the mental processes involved is Paivio's Dual Coding Theory (1991) and Mayer's Multimedia Learning Theory (2005), which suggest that captions improve understanding because they allow the learner to assimilate the verbal information they are receiving in two channels, auditory and visual. This redundancy helps to reduce cognitive overload, aids word recognition, and enables learners to relate speech to the meaning it has (Winke *et al.*, 2013).

Practically, empirical studies have illustrated in many different contexts that captioned videos have been associated with improved listening comprehension, across a range of target language proficiency levels (Montero Perez *et al.*, 2013; Teng, 2019). But captions help understanding, it has been found, and other areas of language learning: incidental vocabulary acquisition, and an awareness of pronunciation (Peters *et al.*, 2016). All these findings strongly suggest that captions are to be seen not merely as subtitles but as pedagogy in themselves, which bridge the gap between the difficulty of input and learner engagement.

2.2. L1 vs. L2 captions: Comparative perspectives

A recent body of research has called into question whether providing subtitles in the learner's first language (L1) or the target language (L2) is of greater benefit. There is much to say for the proposition that L1 subtitles enhance faster semantic access, enabling learners to attend to meaning without being inhibited by inadequate vocabulary or grammatical knowledge (Danan, 2004). This kind of assistance may be especially required by learners who are not well equipped linguistically, or where comprehension rather than awareness of language forms is the main teaching purpose (Montero Perez *et al.*, 2014).

L2 subtitles, on the other hand, are said to produce better or more profound linguistic processing, encouraging learners to map the audible input onto the written L2 forms. In this connection, results from several studies indicate that L2 subtitles produce better results in connection with vocabulary remembering, speech segmentation, and

listening comprehension than do the mere L1 subtitles (Winke *et al.*, 2013; Almusharraf *et al.*, 2024). Furthermore, the partisans of L2 subtitles contend that they afford further opportunities of attending to the authentic use of the language, and thereby further permanent acquisition, not merely ephemeral understanding. However, in fact, the relative superiority of one or other kind of subtitle may depend upon the learner's state of proficiency, the motivation to listen, or the cognitive load borne by the learner. Vanderplank (2016), for example, warned that L2 subtitles could create problems of cognitive overload for low-level learners if too much or too high is the speech rate at which the text is delivered.

However, L1 subtitles may produce a morbid overreliance upon merely translating, preventing the learner from restructuring that auditory information actively from its sense operations. Thus while no commentary serves to provide a format which is universally superior to others, (for the official view upon the current state of commentary in Australia is seen to be complex: for the pithy message in the first page), current studies seem out to show that this is not a matter which makes it possible to generalise until, indeed, it becomes necessary to spend the whole study of learner preferences and influences.

Nevertheless, most of the comparative studies to date have dealt with the media results to be gained, rather than with the subjective perceptions of being to the learner how the different types of subtitles studied were revealed to the learner, rationalizing his preference, say, of subtitles to one kind or another, easy, habitual proneness, or affected educational botany. Thus, the study of learner preference, is commonly made it possible to have been given us with more comprehensive information in reference to subtitles, either surely or otherwise, than would be made at performance eventually, under test, so as to tell the subject.

2.3. Learners' perceptions and affective factors in caption use

The use of captions is more than a cognitive phenomenon because it is highly related to affective variables such as motivation, listening anxiety, and self-efficacy. Li *et al.* (2023) found that students who felt they had more control over captions were more eager to do extensive listening tasks. Likewise, Lee *et al.* (2021) reported that captions led to reduced listening anxiety because of their increased predictability and perceived comprehensibility. These studies would suggest that caption preferences are influenced as much by affect as they are by learning strategy.

Since it is often the case that affective comfort precedes cognitive interest, it is crucial to study how learners perceive various types of captions. A captioning format that lessens stress will probably lead to more frequent volunteer listening, even if that format is objectively not the most "efficient" for acquisition. It is therefore necessary to study self-reported experiences of learners if learner-centered listening instruction is to be successfully generated.

2.4. Caption use in Asian and Vietnamese EFL contexts

Research within East Asian and Middle Eastern EFL contexts shows that in cultures where education is translation-based, learners seem to be much in favor of support from L1 in input, which is difficult (Yuksel & Tanriverdi, 2009; Teng, 2019). Other learners of higher proficiency report favoring captions in L2 so as to give themselves a degree of challenge and to simulate authentic listening contexts. These divergent trends suggest that the anticipations in regard to captions and their use may be determined by their general cultural and educational background.

In the case of Vietnam, research on captions is still limited in extent and mainly directed towards the cognitive effects of captions in regard to comprehension or vocabulary learning rather than learner impressions (e.g., Kien, 2024). To date, there are no systematic studies into how Vietnamese university students view the utility and cognitive effort, and motivational impact of captions in L1 and L2. This "gap" seems to be conducive to the need for perception-based research to provide linguistic and perceptual data for new technology-assisted listening curricula built on caption integration.

2.5. Research gap

While research has widely explored the efficacy of subtitles in improving L2 listening comprehension (especially in Western ESL or high-exposure contexts), there has been a holdback in the current research in exploring how learners in Asian EFL contexts (with limited and exam-oriented exposure to languages) subjectively view the cognitive and emotional trade-offs between L1 and L2 subtitles. Previous studies have focused far more on performance-based outcomes, such as comprehension levels or vocabulary acquisition, rather than on the students' inner reasoning, affective ease, or self-regulated decision-making with respect to the use of subtitles. Where perception is mentioned, most studies have tended to treat all subtitles as a unitary aid to learning rather than a psychological choice between ease (L1) and difficulty (L2).

More importantly, no research is known that has systematically compared L1 to L2 subtitles from the perspective of Vietnamese university learners, despite the widespread use of subtitled video listening in both institutional instruction as well as autonomous learning platforms. The few Vietnamese studies available have mainly touched upon comprehension effects rather than perceived utility, cognitive effort, or motivational impact. As a result, teachers and syllabus designers in Vietnam lack empirical evidence as to how learners experience and justify their subtitles, with the result that decisions are taken on the use of subtitles on the basis of intuition rather than learner-informed considerations. In light of this evidence gap, the current study adopts a perception-based comparative program which emphasizes the students' subjective evaluations and self-reported justifications, thus providing pedagogically-useful evidence for improved subtitles in technology-assisted listening instruction.

3. Research methodology

3.1. Research design

This study followed a sequential explanatory research design (Creswell & Fetters, 2004), using both quantitative and qualitative methods to investigate learners' views about L1 and L2 captions. In the quantitative phase, a survey questionnaire was constructed to determine general patterns in learners' perceived utility and cognitive effort, as well as motivational properties, of the captions in each language. This yielded quantitative data on the patterns of preponderance of the preference for one language or the other in larger populations.

Following the survey, a qualitative phase was conducted through semi-structured interviews with selected subjects. The purpose of this was to clarify and interpret the quantitative findings through the self-reasoning of learners' individual views and emotional responses, in contexts that cannot be satisfactorily explained through closed-ended questions. These two sources of data permitted methodological triangulation, so that the findings were supported by both breadth (survey) and depth (interviews).

3.2. Sample and sampling

The subjects of the study were 130 second-year English major students at Gia Định University (Ho Chi Minh City, Vietnam). All subjects were taking the course Listening and Speaking 2. These subjects were from three intact classes, and thus were considered to have had similar listening experience and curricular requirements. Since it was through the course teachers that access to these classes was possible, the convenience sample of students was used to gather the data taken from the questionnaires, which were given during class times.

During the qualitative phase, there were six students purposively selected from the random subjects for the study. It was felt that it would serve the purpose of balance, and give the data gathered greater significance if a range of tastes in posters or film captions, and the level of enthusiasm of the subjects, were included in the students selected for the interviews. This was felt to be requisite to essay a close and detailed analysis of the subjects' motives for the different perceptive tendencies. All subjects were informed as to the general basis of the study, and consented to its data collection, etc., at a preliminary interview, prior to the collection of any data.

3.3. Research instruments

To investigate Research Question 1 quantitatively, a five-point Likert-scale questionnaire was developed. The instrument consisted of 10 items concerning students' perceptions of L1 and L2 captions along three dimensions: (a) perceived usefulness, (b) cognitive effort, and (c) motivational effect. Each of the items was rated on a scale from 1 = *Strongly Disagree*, 2 = *Disagree*, 3 = *Neutral*, 4 = *Agree*, and 5 = *Strongly Agree*. The items were derived from scales already in use for studying caption perception and listening attitudes (e.g.,

Winke *et al.*, 2013; Teng, 2019) and were examined for clarity and content validity by two experts in language education.

It was thought that this terminate collection of data through a questionnaire means would provide a basis for further data collection through interviews. The qualitative data were obtained through the use of semi-structured interviews involving five open-ended questions. These questions dealt with students' own experience, likes, and reasoning for the choice of L1 or L2 captions, thus adding to and elucidating the results of the survey concerning Research Question 1, and directly dealing with Research Question 2. All the questions of the interview were given in Vietnamese in order to facilitate a natural expression, and students were encouraged to give instances of their own listening experience.

3.4. Data collection procedures

Data collection was implemented in two consecutive phases following a sequential explanatory model. In the first stage, the information was gathered at class times, with consent being sought from the respective course instructors. Participants were made aware of the voluntary nature of the investigation and given an assurance of anonymity prior to completing the questionnaire, which took approximately 10 minutes to complete. Responses were gathered in either hard copy or by means of an electronic link, either via the Web or straight on to a PC, depending on the type of class and the wishes of the instructor.

In the second stage, six individuals were invited for individual semi-structured interviews, selected from survey 1, to ensure variation in caption preference and perception scores. The range of variations in answers permitted possible insight for theme development. Each interview lasted for 10–15 minutes and was carried out in Vietnamese to facilitate a natural mode of expression. The semi-structured interviews were recorded, with the consent of the interviewee, and were subsequently transcribed verbatim for analysis purposes. The two stages of the research were carried out interthematically so that they all took place in the same week to minimize possible bias by retrospective recall and ensure conformity of quantitative and qualitative phases of the investigation.

3.5. Data analysis procedures

The quantitative data generated from the questionnaire were first entered into and held in Microsoft Excel for initial checks, including missing values and inconsistent responses. The cleansed data were subsequently transferred to SPSS for statistical analysis. Before further analysis, Cronbach's alpha coefficients were determined for the questionnaire domains to assess their internal consistency reliability. Descriptive statistics (means and standard deviations) were calculated to show trends of general perception towards the L1 and L2 captions. For the qualitative phase, the recordings of the interviews were transcribed verbatim, and the transcript data analyzed using inductive thematic coding in order to identify recurring themes concerning perceived utility, cognitive effort, and

motivational effects. The themes were then compared with the quantitative data with a view to establishing a convergence or divergence of the two data sets.

4. Results and discussion

4.1. Results

4.1.1. Reliability analysis

Cronbach's alpha coefficients were calculated on the three respective subscales in order to assess the internal consistency of the questionnaire. In this case, they indicated satisfactory levels of reliability, as presented in Table 4.1, for the three subscales: that for the perceived usefulness was .82, cognitive effort .78, and motivational effects .85, all values exceeding the generally accepted limit of .70 (Hair *et al.*, 2019).

Table 4.1: Reliability of subscales (N = 130)

Subscales	Number of Items	Cronbach's Alpha
Perceived Usefulness	4	.82
Cognitive Effort	3	.78
Motivational Impact	3	.85
Total	10	.82

4.1.2. Vietnamese university students' perceptions of L1 captions compared with L2 captions

Before moving on to interpretations at the subscale level, a descriptive overview of perceptions of all 10 items was provided in order to clarify the general attitudinal landscape toward these captions (L1 and L2) in common usage. These data are represented in Table 4.2, and indicate that the mean scores of all items fluctuated from 3.30 to 3.60, thus indicating moderately positive perceptions in general terms without extreme polarization toward either type of caption.

Students differentiated L1 and L2 captions on the four usefulness-oriented items. Captions in L1 were more useful for immediate comprehension, while L2 captions were seen as more helpful for long-term language development. This is observed in the mean scores of item one (M=3.53; SD=.799) and item three (M=3.39; SD=.857), where learners noted that Vietnamese captions were helpful for understanding global meaning and for maintaining continuity over the rest of the audio when the audio became too fast or complicated. Meanwhile, items 2 (M=3.53; SD=.837) and 4 (M=3.39; SD=.902) suggest that English captions received credit for helping with vocabulary retention and deeper processing, even if that came with slower development. This distribution suggests students did not evaluate the superiority of one captioning format in absolute terms, but rather assigned them different functions: L1 for stability and L2 for development. This functional distinction indicates that learners implicitly and strategically managed caption selection for their listening goals, a behavior consistent with metacognitive regulation theories in L2 learning.

Table 4.2: Vietnamese university students' perceptions of L1 captions compared with L2 captions

No.	Subscales	Items	N = 130	
			M	S.D.
1	Perceived Usefulness	L1 captions help me understand video content more clearly than L2 captions.	3.53	.799
2		L2 captions help me learn new vocabulary better than L1 captions.	3.53	.837
3		L1 captions make it easier for me to follow the main ideas, compared to L2 captions.	3.39	.857
4		L2 captions contribute more to my long-term learning than L1 captions.	3.39	.902
5	Cognitive Effort	I need less mental effort when using L1 captions than when using L2 captions.	3.42	.756
6		L2 captions sometimes overload me more than L1 captions.	3.39	.751
7		Switching attention between listening and reading is easier with L1 captions than with L2 captions.	3.42	.755
8	Motivational Impact	I feel more confident when watching videos with L1 captions than with L2 captions.	3.35	.776
9		L2 captions motivate me to challenge myself more than L1 captions.	3.56	.747
10		I am more likely to continue using L2 captions in future listening practice than L1 captions.	3.41	.842

Notes: M = Mean; S.D. = Standard Deviation.

A student made this differentiation in the interview most clearly: *"If I only want to enjoy or understand it quickly, I choose Vietnamese captions. But when I have a mood for serious learning, I have English captions on. These captions mean I have to realize how words are used. This is slower in a way, but I feel that I am getting better"* (student 5). This quote embodies the dichotomy between efficiency and proficiency, which stands in contrast to the students' decision-making process. Rather than being passive consumers, it appears they implement self-regulated modality control in managing the captions, passing consciously from comfort-oriented (L1) to effort-oriented (L2) modes. The language *"I feel like I am learning seriously"* indicates that intentionality and learner identity are attached to caption selection — L2 captions are not simply pedagogical instruments, but features of agency on the part of learners when self-demarcating as active L2 learners rather than strictly passive consumers. This inference of learner control serves to augment the view that caption selection is not simply cognitive, but motivationally symbolic, and that pedagogical decisions regarding caption input must acknowledge both functional and emotional aspects of the usefulness dimension.

As pertains to the second subscale :(Cognitive effort) Perceptions of cognitive effort revealed a functional distinction in clear terms between L1 and L2 captions. However, unlike usefulness, which was cast in terms of a compromise, effort appeared to be judged in more emotional terms, usually in the context of stress management rather

than learning strategy. The descriptive statistics of item 5 ($M=3.42$; $SD=.756$) and item 7 ($M=3.42$; $SD=.755$) depict that respondents experienced little in the way of cognitive strain where L1 captions were used. Thus, it follows that captioning in Vietnamese served a stabilizing function cognitively, allowing the students to process incoming auditory information without losing the run of things. On the other hand, item 6 ($M=3.39$; $SD=.751$) illustrates that, at times, L2 captions were overwhelming, particularly when the learners encountered difficulties in reconciling the dual tasks of both reading and listening simultaneously. Students appeared not to reject L2 captions, although they seemed to reason with the difficulty, tolerating the mental strain in circumstances when it possibly accrued to them a benefit in perceived gain in the long run. It follows from this that the concept of effort was not perceived as purely negative but contextualized into a threshold of manageability personally conceived.

One of the interview subjects described this delicate nicety in obvious terms: *"When I use English captions, my mind seems to work too hard. I lose it sometimes as I am reading it and listening to it, but if I slow the picture down or re-run it that is okay again. That is tiring, but it is a nice tiring"* (Student 4). It is opportune, too, that this evidence is salutary, not only because it supports item 6 as an expression of situational effort instead of as an express rejection but in fact more so because it makes a significant distinction between "unproductive effort" and "acceptable effort." The student's use of the term *"an acceptable tiring"* suggests that students are willing to suffer cognitive load when it is within their control. This point indicates a little further that is to say, that in the case of L1 captions, these are conceived of as being cognitively "safe". L2 captions, on the other hand, are conceived of as being cognitively *"taxing but worth the trouble"* when agency is supplied. The fact emerges, reluctantly, that caption design must not only contemplate an effort calibration but must also have a certain autonomous scope for the user, since learners begin to show they are willing to absorb the cognitive impact if they have a tendency throughout as to tempo and density.

While usefulness and cognitive effort expressed a negotiation between ease and difficulty, the motivational findings offered a more affective differentiation: L1 subtitles provided emotional comfort; L2 subtitles served as motivational fuel. The descriptive statistics describe that L1 subtitles related to the motivational construct of confidence-building (Item 8; $M=3.35$; $SD=.776$), signifying that they offered a sense of comfort and safety net that protected learners against disengagement from study, now dealing with the unfamiliar. However, item 9 demonstrates that the L2 subtitles were superior in stimulating challenge-seeking behavior (Item 9; $M= 3.56$) and in orientation to future commitment (Item 10; $M= 3.41$). This dualism leads to the conclusion that learners did not merely distinguish the subtitles as easy or difficult, but rather as *"comfort mode"* or *"growth mode."* The subtitles, in this light, became identity markers: selection of the L1 runs to self-protection while choice of the L2 was judged to be in the area of intention to self-upgrade.

This distinction was pointed out clearly by one of the interviewees: *"Vietnamese subtitles make me feel relaxed, as if I cannot fail. But English subtitles make me feel that I am*

learning something. When I complete a video that has English subtitles, I feel proud of myself" (Student 6). It gives one the sense of a psychological key, that motivation is to be understood not so much in terms of enjoyment as of self-validation. Instead of being forced by external pressures, the students tend to assess L2 subtitles as a means of achievement and self-growth even when their comprehension is not perfect. The sentence, *"I feel proud of myself,"* tends to indicate that L2 subtitles carry a symbolic load as an indication of persistence rather than mastery. The problem with the pedagogy for providing the subtitles should not be simply looked upon as a binary tool for greater access, but rather as a motivational architecture, where L1 subtitles lower the withdrawal risks and L2 subtitles fan the flames of aspirational engagement in the second language. When considered together, the descriptive findings emerge to suggest a dual motivational dynamic, with L1 captions appearing to anchor learners within feelings of safety and fluency, while L2 captions appear to activate aspirations for improvement and autonomy. In order to better understand how students position themselves in this spectrum, the next section examines patterns of caption preferences, focusing on whether learners tend to affiliate with one caption type at the expense of the other or vice versa, or to switch strategically between the two options.

4.1.3. Students' preferences for either L1 or L2 captions when engaging with video-based listening materials

A comparison of preference data paints a picture of a split: roughly 57 % of the participants leaned toward L1 captions while 43 % chose L2 captions. This balance does not signal a win for either side; rather, it hints that students' choices echo distinct learning philosophies. Those who favored L1 captions generally treated listening as an exercise valuing the progression of ideas and the continuity of the message, above all. By contrast, the cohort that favored L2 subtitles saw listening as a sort of laboratory, for enrichment and were ready to tolerate short-lived discomfort for the promise of longer-term payoff. Students who leaned toward L1 subtitles consistently described their choice in terms of steadiness and emotional comfort. For them, Vietnamese captions functioned as a kind of safety net warding off irritation and keeping focus intact during listening stretches. As one participant put it succinctly: *"When the captions, in Vietnamese, I don't panic about missing words. I can. Just follow the story without worrying about whether I understand every detail"* (Student 2). What this quotation points to is a sort of built-in anxiety-regulation system: L1 subtitles are not a crutch for comprehension; they serve as a buffer that keeps the tide of information from spilling over into demotivation. Rather than marking laziness or avoidance, the habit of relying on them reads like a self-preservation tactic—a way to stay hooked on the audio without letting confidence collapse.

On the one hand, the students who favored English subtitles framed their choice in terms of identity, then convenience. To them, English captions felt like a commitment of passive consumption. When they hit moments where they missed something, they did not see it as a failure. As evidence of a productive struggle. As one participant put it, *"English captions make me feel like I'm actually pushing myself. Even if I don't catch everything,*

I know my brain is working harder in a way" (Student 6). This mindset shows that L2 caption preference is more motivational than functional. Learners aren't simply opting for the choice—they are aligning themselves with a self-image of discipline, resilience, and a future-oriented outlook. In this sense, the caption preference becomes an expression of learning identity rather than a processing style.

Taken together, the split between L1-leaning "*stability-oriented listeners*" and L2-leaning "*challenge-oriented listeners*" shows that captions are more than an add-on - they act as a psychological positioning tool. Of mandating a caption style, educators might benefit from flexible caption policies that accommodate both listener profiles. The ramifications of these identity pathways will be unpacked in the chapter.

4.2. Discussion

The results of this study support the well-documented functional dichotomy between L1 and L2 subtitles that has been recorded in prior studies (e.g., Montero Perez *et al.*, 2014; Winke *et al.*, 2013), yet the parameters of such studies have been expanded by indicating how learners view subtitles not so much as a means of aid to comprehension, but as identity constructs linked to different learning objectives. This is consistent with Vanderplank's (2016) assertion that L1 supports continuity of meaning, while L2 predominantly encourages depth of encoding, as Vietnamese university students in this study clearly assigned different psychological roles to each form of subtitle. L1 was assigned the status of a control mechanism and L2 that of a growth mechanism. Contrary to showing a preference for a particular form of subtitle universally, learners showed considerable adaptability to the different forms, switching between L1 and L2 subtitles according to whether their short-term objectives were pleasure or conscious practice. This reflects Paivio's Dual Coding theory in practice. Subtitles were not perceived as static elements of visual support, but as metacognitive levers that could control processing loads.

An original insight gained from the present study is that the preference for subtitles was not determined by cognitive efficiency alone, but rather by the emotional factors of manageable effort and self-reflexiveness. The interviews indicated that effort, with L2 subtitles, was often the result of feeling "good tired," which is to say it was felt that some cognitive drain was beneficial, provided that it could be maintained within manageable areas of control. This is again consistent with Li's (2022), who noted that not so much the actual gain in educational effect by subtitles as the subjective autonomy felt to exist was of greater benefit from the human standpoint. Also, the Lee *et al.* (2021) argument that the anxiety for learners caused by the use of subtitles is alleviated was confirmed with L1 subtitles only; the feeling likely increased with the availability of the L2 subtitles, but simultaneously increased the efficacy felt. The statement "I feel good about myself when I finish off with English subtitles" could be equated to the fact that L2 subtitles are trophies to be obtained, as opposed to comforts to be obtained. The dividing line here, then, is not one of comfort versus difficulty, but of safe mode versus growth-inducing period.

Results such as the above suggest that the decisions taken pedagogically with regard to the use of subtitles must change from the old binary one of “which form of subtitle is better” to one of dual-pathway disposition, with recognition of the acceptance of two different identity paths taken by learners. There is seen to be no satisfactory basis for educators to enforce either solely the use of L2 subtitles in the interest of rigor, or else to use L1 subtitles on the premise of individual convenience. Instead, subtitles must come from a position of modification into flexible vehicles for engagement, which can allow learners to use L1 subtitles at the outstart for a stage of confidence-building, conditioning them to the task, and moving to the use of L2 subtitles when they feel more confident towards challenge-based learning and divergence of the brain systems used. Those engaged in the design of platforms and preparation of curricula might consider open-ended for systems where switchings of subtitles around could be made at will, or by possibility aging from L1 to L2 modes. Ultimately, the use of subtitles should not be considered as a deep-seated technical facet, but rather a self-controlled strategy which gives learners the opportunities to negotiate satisfactorily the twin positions of comfort and aspiration.

4.3. Limitations of the study

Even though the study offers insights, it carries a few caveats that should be taken into account when the findings are interpreted. First, the participants were exclusively second-year students from a single Vietnamese university, which may curb the extent to which the results can be generalized to learners of other proficiency levels or disciplines. Students in majors or in secondary-school settings might perceive captions differently, perhaps because they had less exposure to autonomous listening or faced different levels of academic pressure. The quantitative instrument, in turn, rested on self-reported perceptions rather than on observable behavioral indices, thereby introducing the possibility that learners' declared preferences could diverge from their genuine caption-use behavior in authentic, real-time listening scenarios.

Moreover, even though the mixed-methods design offered a deep view of the qualitative portion, it rested on just six interviewees - hardly enough to capture the full spectrum of learners' motivational orientations or identity-based reasoning. Future studies might add tracking or experimentally tweak caption-switching to see how learners' preferences shift with changes in difficulty level, task type, or emotional state. Cross-institutional comparative work - or a side-by-side look at caption choices in a brick-and-mortar classroom versus a digital learning environment - might untangle whether the tilt toward captions springs more from the instructional setting than from an individual's own learning ethos.

5. Conclusion

This study explored how Vietnamese university students perceive first-language (L1) and second-language (L2) captions in video-based listening tasks, showing that they did

not view subtitles as a binary of superiority but rather as context-dependent tools that fulfill distinct cognitive and motivational roles. L1 captions were prized for the reassurance and processing stability they afford, while L2 captions were seen as drivers of long-term linguistic growth and sources of self-validation. Of committing to a single caption style, learners purposefully switched between a "comfort mode" and a "growth mode" to match their listening goals, suggesting that captions functioned more as a metacognitive regulation tool than as passive support. These results point to the need for designs that offer flexible caption options and scaffolded pathways for transitioning between L1 and L2 assistance. Future work should dig into how caption-switching varies across learners with proficiency levels and across a range of task types, aiming to craft adaptive captioning frameworks that honor learners' identities, empower their agency, and keep pace with their evolving linguistic readiness.

Acknowledgements

The author wishes to convey thanks to Gia Dinh University in Vietnam for the institutional funding that enabled this research. Particular gratitude goes to the student participants for their time and valuable input. The author also acknowledges the help of colleagues and lecturers in gathering and analyzing data, as well as the academic peers whose suggestions helped improve the study's quality. Finally, a few sections of this manuscript were polished with AI-powered language tools (i.e., ChatGPT, Grammarly) to improve clarity and coherence. The author, however, maintains ownership of every idea, logical argument, and interpretation presented.

Creative Commons License Statement

This research work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc-nd/4.0>. To view the complete legal code, visit <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.en>. Under the terms of this license, members of the community may copy, distribute, and transmit the article, provided that proper, prominent, and unambiguous attribution is given to the authors, and the material is not used for commercial purposes or modified in any way. Reuse is only allowed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Conflict of interest statement

The author declares no conflicts of interest.

About the author

Nguyen Thi Bach Lien is currently teaching at Gia Dinh University as a part-time lecturer. She graduated from Can Tho University with a Bachelor's Degree in English and also holds a Master's Degree in TESOL. She has had over 10 years of experience as an English teacher in both private schools and foreign language centers. With a great passion and

flexible style of teaching, as well as being talented in using different teaching methods, she can cater for all students and abilities by physical demonstrations and visual presentations.

References

- Almusharraf, A., Mahdi, H. S., Al-Nofaie, H., Ghobain, E., & Aljasser, A. (2024). Video captioning and subtitles in second language listening comprehension: Fast-paced versus slow-paced speakers. *Journal of psycholinguistic research*, 53(2), 29. <https://doi.org/10.1007/s10936-024-10070-z>
- Creswell, J. W., Fetters, M. D., & Ivankova, N. V. (2004). Designing a mixed methods study in primary care. *The Annals of Family Medicine*, 2(1), 7-12. <https://doi.org/10.1370/afm.104>
- Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta*, 49(1), 67-77. <https://doi.org/10.7202/009021ar>
- Field, J. (2008). *Listening in the language classroom*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511575945>
- Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European business review*, 31(1), 2-24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Kien, T. T. (2024). The effectiveness of using visual communication in teaching vocabulary to non-English majors students at the Foreign Languages Center of Nguyen Tat Thanh University. *Journal of Science and Technology*, 7(3). <https://doi.org/10.55401/c1kpm641>
- Lee, P. J., Liu, Y. T., & Tseng, W. T. (2021). One size fits all? In search of the desirable caption display for second language learners with different caption reliance in listening comprehension. *Language Teaching Research*, 25(3), 400-430. <https://doi.org/10.1177/1362168819856451>
- Li, K., Qin, X., Ji, S., & Zou, L. (2023). The effects of general listening anxiety and listening test anxiety on self-perceived listening performance among Chinese English learners. *Psychology Research and Behavior Management*, 3541-3560. <https://doi.org/10.2147/PRBM.S422030>
- Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511816819>
- Montero Perez, M., Van Den Noortgate, W., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, 41(3), 720-739. <https://doi.org/10.1016/j.system.2013.07.013>
- Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, 45(3), 255-287. <https://doi.org/10.1037/h0084295>

- Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134–148. <https://doi.org/10.1016/j.system.2016.10.002>
- Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Pearson Education. <https://doi.org/10.4324/9781315833705>
- Teng, F. (2019). The effects of captioned videos on L2 listening comprehension. *ELT Journal*, 73(2), 140–150. <https://doi.org/10.1093/elt/ccy054>
- Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge. <https://doi.org/10.4324/9780203843376>
- Vanderplank, R. (2016). *Captioned media in foreign language learning and teaching: Subtitles for the deaf and hard-of-hearing as tools for language learning*. Palgrave Macmillan. <https://doi.org/10.1057/9781137500458>
- Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, 97(1), 254–275. <https://doi.org/10.1111/j.1540-4781.2013.01435.x>
- Yuksel, D., & Tanriverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *The Turkish Online Journal of Educational Technology*, 8(2), 48–54. Retrieved from <https://files.eric.ed.gov/fulltext/ED505936.pdf>.