The International Journal of Language Studies
2025, Vol. 3, No. 1, pp. 12–28
Journal page: https://ijlangstudies.org/index.php/home
DOI: https://doi.org/10.60087/ijls.v3.n1.002
ISSN 3078-2244

Research Article

CULTURAL CONCEPTUALISATIONS IN VISUAL ARTIFACTS OF VIETNAMESE EFL MATERIALS: A CULTURAL LINGUISTICS ANALYSIS OF AN ACADEMIC ENGLISH TEXTBOOK

Nguyen Minh Nam Anh [1]

[1] Department of Foreign Languages, Foreign Trade University – HCMC Campus, Vietnam

Abstract

This study investigates the role of visual artifacts in an academic English listening and speaking textbook commonly used in Vietnamese EFL contexts, employing the framework of cultural linguistics as proposed by Sharifian (2017). Drawing on the concepts of cultural schemas and cultural conceptualisations, the analysis examines how visuals, such as representations of urbanisation in architecture-themed units or environmental degradation in ecology-focused sections, primarily reflect Western-centric frames of reference, with limited representation of Vietnamese cultural categories such as collectivist harmony or ecological balance, rooted in Confucian and agrarian traditions. Utilising multimodal discourse analysis (Kress & van Leeuwen, 2006) on selected visuals and qualitative surveys with 37 Vietnamese university students, the research identifies intercultural mismatches that hinder metacultural competence, potentially leading to pragmatic failures in cross-cultural communication. Findings suggest that supplementing these visuals with localised elements, such as depictions of Vietnamese communal festivals or rural landscapes, can enhance learners' ability to negotiate distributed cognition across linguistic boundaries. Implications for ELT materials design underscore the need for glocalised approaches in non-native speaker contexts, aligning with recent scholarship on visual semiotics in Asian EFL textbooks (e.g., Chen, 2010).

Keywords

Cultural conceptualisations; Cultural linguistics; Visual artifacts; Metacultural competence.

1. Introduction

In the field of English as a Foreign Language (EFL) pedagogy, visual artifacts embedded within instructional materials serve as potent semiotic resources that extend beyond mere illustration, encoding complex cultural conceptualisations that shape learners' interpretive frameworks (Sharifian, 2017). These artifacts, ranging from photographs and diagrams to infographics, function as multimodal ensembles where meaning emerges from the interplay of visual and textual modes, often reflecting dominant ideologies that may alienate non-Western learners (Kress & van Leeuwen, 2006). In Vietnamese EFL contexts, where English language teaching (ELT) materials are

*Corresponding author: Nguyen Minh Nam Anh

Email addresses:

nguyenminhnamanh.hcmc@ftu.edu.vn (Nguyen Minh Nam Anh)

Received: 05/02/2026; **Accepted:** 11/03/2026; **Published:** 26/04/2026

Copyright: © The Author(s), 2026. Published by JKLST. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

predominantly sourced from Anglo-American publishers, textbook visuals often privilege native-speaker or Western cultural frameworks while giving limited space to local cultural representations (Vuong et al., 2025). This study undertakes a cultural linguistics analysis of the academic English textbook *Prism 3: Listening and Speaking* (Lansford et al., 2017), a widely adopted resource in Vietnamese tertiary institutions, to interrogate how its visual components in units on architecture and the environment create intercultural frictions that, if left unaddressed, may impede the development of metacultural competence; however, these same frictions can become pedagogical resources when explicitly engaged – the ability to negotiate culturally distributed cognition across linguistic boundaries (Sharifian, 2011).

The widespread use of EFL materials in non-native speaker settings has long faced criticism for their ethnocentric bias, in which visual semiotics unintentionally prioritize Western epistemologies, thereby undermining local cultural schemas (Canagarajah, 1999). For example, in *Prism 3*, the visuals in Unit 5 (Architecture) portray skyscrapers and urban skylines as symbols of progress and personal success, aligning with capitalist ideologies centered on transformation and innovation (Lansford et al., 2017, pp. 102-123). These depictions evoke cultural conceptualizations of urbanization as a straightforward path to modernity, frequently relying on schemas of technological superiority that echo Western development narratives. Likewise, in Unit 4 (The Environment), images of cloning laboratories and endangered species highlight technocratic approaches to habitat loss, framing environmental problems through the perspective of scientific intervention rather than comprehensive balance (Lansford et al., 2017, pp. 80-101). Although these visuals incorporate ideologies that contrast with Vietnamese cultural frameworks, such as the Confucian focus on collective harmony and agrarian ecological equilibrium, where nature is viewed not as a resource for engineering but as an interconnected system supporting community life (Tran, 2019).

At the same time, it is essential to recognize that Prism 3 is designed as an Academic English textbook. Its choice of topics – cloning, skyscrapers, global climate change – reflects intentional efforts to prepare learners for participation in international scientific and economic discourse. These are not arbitrary Western impositions but legitimate academic content areas that Vietnamese students need to navigate as globally mobile professionals. The critique offered here is not that such content should be removed, but rather that it should be supplemented and critically examined so that learners can engage with global concepts without losing access to their own cultural resources.

The cultural linguistics framework developed by Sharifian (2017) offers a powerful lens for analyzing these issues. It argues that language and culture are deeply connected through shared concepts, mental processes that appear in both language and visual signs. According to this view, cultural schemas are the shared mental frameworks a community uses to understand the world, while cultural conceptualisations are how these frameworks are actively expressed in communication, including through images. This perspective is effectively combined with multimodal discourse analysis (MDA) (Kress & van Leeuwen, 2006), which studies how visuals create meaning. This approach identifies three key functions: representational (what is shown, like using skyscrapers to symbolize progress), interactional (how the image engages the viewer, such as a high-angle shot creating a sense of authority), and compositional (how elements are arranged for effect, like placing modern and traditional buildings side-by-side to emphasize contrast). In English language learning, when textbook visuals are designed using these functions from a predominantly Western viewpoint, they can lead to pragmatic failures – breakdowns in intended meaning. This happens because the images may not align with the distributed cognition, or the shared, culturally specific ways of thinking, of Vietnamese learners (Sharifian, 2011).

The Vietnamese context of English language education provides a clear case study of these tensions. The global dominance of English and the pressure for internationalization in higher education have made internationally published textbooks a common, and often unquestioned, choice (Pennycook, 2017). While series such as *Prism* from Cambridge University Press are valued for integrating academic skills with contemporary global themes, a cultural linguistics analysis reveals that their visual discourse frequently privileges Western cultural schemas. For instance, imagery framing environmental issues may present technological solutions as paramount, inadvertently marginalizing place-based epistemologies. This contrasts with Vietnamese cultural conceptualizations, where ecological balance is traditionally viewed as an interdependent system tied to social harmony and community stewardship (Tran, 2019). This marginalization of local schemas inhibits the development of learners' metacultural competence – their ability to consciously navigate between different cultural conceptual systems. Ultimately, this dynamic risks aligning with critiques of linguistic imperialism, where the global spread of English promotes the cultures and knowledge systems embedded in its teaching materials as normative (Phillipson, 1992). This issue is not isolated. Research in similar Asian EFL contexts confirms that a reliance on Western-centric content can create intercultural mismatches, hindering effective cross-cultural communication by failing to engage with learners' own cultural cognitive landscapes (Setyono & Widodo, 2019).

2. Literature Review

This review synthesizes three interconnected strands of scholarship essential to understanding the role of visual artifacts in EFL materials through a Cultural Linguistics lens: (1) the theoretical foundations of Cultural Linguistics; (2) the semiotic theory and analysis of visual artifacts; and (3) the ideological and political critiques of global EFL textbooks in Asian and Vietnamese contexts.

2.1. Cultural Linguistics: Foundations in conceptualisations and competence

Emerging at the intersection of cognitive linguistics and anthropological linguistics, Cultural Linguistics (Sharifian, 2011, 2017) provides a robust theoretical framework for examining the intrinsic relationship between language, culture, and cognition. It posits that human language is fundamentally underpinned by cultural conceptualisations – shared, collective cognitive representations that emerge from the interactions between members of a cultural group over time. These conceptualisations manifest primarily through:

(1) Cultural schemas: Organized, default knowledge structures about objects, events, and situations (e.g., schemas for "progress", "success", or "ecological balance").

(2) Cultural categories: Culturally constructed ways of classifying experiences and the world (e.g., categories of "modern vs. traditional architecture").

(3) Cultural metaphors: Figurative mappings from a source to a target domain that are conventionalized within a cultural group (e.g., "NATURE IS A MACHINE" vs. "NATURE IS A NETWORK").

These conceptualisations constitute a distributed cultural cognition – a form of cognition that is shared across, and exists between, the minds of individuals belonging to a cultural group (Sharifian, 2017). From this perspective, communication is essentially an act of negotiating these distributed conceptualisations.

A key pedagogical imperative arising from this framework is the development of metacultural competence (Sharifian, 2013). This goes beyond traditional notions of intercultural competence by focusing on learners' strategic awareness of cultural conceptualisations themselves. It involves the ability to consciously recognize, interpret, negotiate, and employ different cultural conceptual systems in communication. In EFL contexts, a lack of such competence – often resulting from materials that present a monolithic cultural worldview – can lead to pragmatic failure, where linguistic accuracy is achieved but the intended cultural meaning is misconstrued (Thomas, 1983).

2.2. Visual Artifacts as semiotic resources: Theory and analysis

2.2.1. The role of Visual Artifacts in meaning-making

In contemporary EFL pedagogy, visual artifacts including photographs, illustrations, infographics, and diagrams are recognized as constitutive elements of the curriculum, not merely decorative supplements (Kress, 2010). They function as potent semiotic resources, systems of meaning-making with their own grammars and affordances that operate in constant interplay with textual language (Jewitt, 2009). Critically, visual artifacts are ideologically encoded; they are always constructed from a particular perspective, making selective representations that promote specific ways of knowing and seeing the world while marginalizing others (Rose, 2016). Therefore, in globally distributed EFL textbooks, images become primary carriers of implicit cultural conceptualisations, subtly reinforcing the cognitive frameworks of their producers (Chen, 2010).

2.2.2. A framework for deconstruction: Multimodal Discourse Analysis

To systematically analyze how visual artifacts construct meaning, this study employs Multimodal Discourse Analysis (MDA) as developed by Kress and van Leeuwen (2001, 2006). Rooted in Hallidayan systemic-functional linguistics, MDA applies the concept of metafunctions to visual communication, positing that every image simultaneously performs three kinds of work:

(1) Representational/Ideational meaning: This addresses what is depicted. It analyzes how images represent participants, actions, events, and conceptual structures through narrative processes (e.g., vectors showing action) and conceptual processes (e.g., symbolic attributes, classificational trees). An image of a lone scientist in a lab, for instance, may conceptually symbolize "technological salvation".

(2) Interactive/Interpersonal Meaning: This addresses the relationship the image establishes with the viewer. Key analytical tools include contact (direct gaze demanding engagement), social distance (close-up suggesting intimacy), perspective (subjective angle inviting identification, objective angle asserting authority), and modality (the coded level of realism or truth claim).

(3) Compositional/Textual Meaning: This addresses how the elements are integrated into a coherent whole. It examines information value (the significance conveyed by placement – center/margin, left/right), salience (how elements attract attention via color, size, focus), and framing (the use of lines or space to connect or disconnect elements).

This tripartite framework allows for a rigorous, replicable analysis of how the formal properties of a textbook image

materialize specific cultural schemas and ideologies, making visible the potential points of alignment or clash with a learner's cognitive environment.

2.3. The cultural politics of global textbooks in Vietnam

The widespread use of internationally-published textbooks in non-Western countries is closely linked to the global influence of English. Many critical scholars see this as a form of linguistic imperialism, where Western values and ways of thinking are presented as neutral and universal (Phillipson, 1992; Pennycook, 2017; Canagarajah, 1999). This becomes even stronger with the ideology of native-speakerism, which treats native speakers as the ideal model for learners (Holliday, 2006). As a result, textbooks may unintentionally marginalize local knowledge and cultural beliefs (Bourdieu, 1991).

Research from Asian EFL contexts supports this concern. Studies in Indonesia (Setyono & Widodo, 2019), Taiwan (Chen, 2010), and Japan show that textbook images often portray Western urban lifestyles that are consumerist and individualistic. Such images promote a view of material progress that can make learners feel disconnected from their own community-oriented values. Vietnam is a clear example of this tension. Due to globalization, textbooks like Prism have become a common choice in many universities (Vuong et al., 2025). However, their cultural messages often conflict with traditional Vietnamese values, such as Confucian collectivism, community harmony, and a balanced relationship with nature (Trần, 2019). In response, scholars have called for critical pedagogy and glocalisation strategies, which involve adding local content to global materials and helping learners think more critically about the hidden messages in those materials (Crookes, 2013; Robertson, 1995).

3. Methodology

3.1. Research design

The research design integrates two complementary components to address the study's objectives:

(1) Textual-visual analysis: A systematic multimodal discourse analysis of selected visual artifacts from Prism 3.

(2) Qualitative survey: An open-ended questionnaire administered to Vietnamese university students to elicit their interpretations of and responses to the identified visuals.

This dual design enables triangulation of data sources, allowing the researcher to compare the cultural conceptualisations encoded in the visuals (as identified through MDA) with the actual interpretive frameworks activated in Vietnamese learners (as captured through surveys). Such triangulation enhances the depth and credibility of the findings (Flick, 2018).

3.2. Data sources

3.2.1. Textbook selection

The primary data source is *Prism 3: Listening and Speaking* (Lansford et al., 2017), published by Cambridge University Press. This textbook was selected for three reasons:

(1) Widespread adoption: It is extensively used in English preparatory programs and undergraduate courses at Vietnamese tertiary institutions, including the researcher's own university, making it a relevant and ecologically valid choice.

(2) Academic focus: As an academic English textbook designed for upper-intermediate to advanced learners, it targets students preparing for academic study in English-medium contexts, where metacultural competence is particularly critical.

(3) Thematic relevance: The textbook contains units on architecture (Unit 5) and the environment (Unit 4), themes that are explicitly mentioned in the study's introduction as sites of potential cultural contestation between Western and Vietnamese worldviews.

3.2.2. Visual artifact selection

From these two units, a purposive sample of visual artifacts was selected based on the following criteria (adopted from Serafini, 2014): (1) Salience: Images that occupy prominent positions within the unit (e.g., opening spreads, large-format images, images accompanying core tasks); (2) Thematic centrality: Images directly related to the unit's core concepts (urbanisation, architectural innovation, environmental degradation, technological solutions); (3) Analytical richness: Images containing sufficient semiotic density to warrant detailed multimodal analysis (e.g., multiple participants, complex compositions, symbolic elements).

A total of 14 visual artifacts were selected for in-depth analysis: 6 from Unit 4 (The Environment) and 6 from Unit 5 (Architecture). This sample size is consistent with similar qualitative multimodal studies of textbook imagery (Chen, 2010; Setyono & Widodo, 2019).

Due to copyright restrictions, this study does not reproduce any original images from Prism 3: Listening and Speaking (Lansford et al., 2017). Instead, each visual artifact is identified by a descriptive label, its page number, and a brief description of its content (e.g., "Image 2a, p. 82: A banteng calf looking directly at the camera in a research facility"). This approach allows for detailed multimodal analysis while fully respecting the publisher's intellectual property rights.

3.3. Participants

The participant cohort comprised 37 Vietnamese university students enrolled in an English subject at Foreign Trade University – Ho Chi Minh City Campus. Participants were recruited through convenience sampling, as the researcher had direct access to this population. However, efforts were made to ensure diversity within the sample: 19 females and 18 males from various disciplines (International Business, Economics, Finance) to avoid disciplinary bias. All students were at B2 level (upper-intermediate) or above, as determined by institutional placement tests, ensuring they could comprehend the textbook content and articulate their interpretations in either English or Vietnamese. Participation was voluntary, and students were assured that non-participation would not affect their course grades. Informed consent was obtained from all participants prior to data collection.

It is important to acknowledge that this sample is drawn from an urban, elite university (Foreign Trade University – HCMC Campus). These students have greater exposure to globalized media, international curricula, and English-dominated professional aspirations than the average Vietnamese EFL learner. Consequently, their cultural schemas are already highly hybridized and may not represent the broader Vietnamese learner population, particularly those in rural areas or less internationally oriented institutions. This study therefore treats the sample as illustrative of a specific learner segment rather than as a monolithic baseline for Vietnamese culture.

3.4. Data collection instruments

To elicit participants' interpretations of the selected visuals, an open-ended questionnaire was developed. The questionnaire was designed to probe the three dimensions of meaning identified in MDA (representational, interactive, compositional) while remaining accessible to student participants.

The questionnaire consisted of three sections:

Section A: Demographic information (age, gender, major, years of English study).

Section B: Visual interpretation tasks. For each of the 14 selected images, participants were asked the following questions: (1) "What do you see in this image? Please describe it in detail."; (2) "What do you think is the main message or idea of this image?"; (3) "How does this image make you feel? Does it feel familiar or unfamiliar to you? Please explain."; (4) "Do you think this image relates to Vietnamese culture or your own life experiences? Why or why not?"

Section C: Reflective questions (administered after all images were viewed): "Overall, do you feel the images in this textbook represent your own cultural experiences? Why?"; (2) "Can you think of any Vietnamese images, places, or scenes that would better represent the themes of 'architecture' or 'environment'?"

The questionnaire was piloted with three students (not part of the main sample) to check for clarity, comprehensibility, and appropriate length. Minor wording adjustments were made based on pilot feedback. Participants could respond in either English or Vietnamese to ensure they could fully express their thoughts without language constraint. Responses in Vietnamese were later translated into English by the researcher and verified by a second bilingual researcher.

3.5. Analytical framework and procedures

Data analysis proceeded in two interconnected phases, corresponding to the two data sources.

3.5.1. Phase 1: MDA of Visual artifacts

The 14 selected images were analyzed using Kress and van Leeuwen's (2006) grammar of visual design. A structured analytical framework was developed, operationalizing the three metafunctions:

Table 3.5.1. *Analytical framework for multimodal discourse analysis of visual artifacts (Adapted from Kress & van Leeuwen, 2006)*

Metafunction	Analytical focus	Key questions
Representational	Narrative vs. conceptual processes; Participants and circumstances; Symbolic attributes	What is happening? Who/what is depicted? What are they doing? What concepts are being symbolized?
Interactive	Contact (demand/offer); Social distance (intimate/personal/social/public); Perspective (subjective/objective, horizontal/vertical angle); Modality (color, contextualization, representation)	How does the image engage the viewer? What power relationship is constructed? How "real" does the image appear?
Compositional	Information value (given/new, ideal/real, center/margin); Salience (size, color, focus, foregrounding); Framing (connection/segregation)	How are elements arranged? What is made most prominent? What is connected or separated?

For each image, detailed analytical notes were compiled,

and a summary table was created capturing the key features of each metafunction. These analyses were then synthesized to identify recurring cultural schemas (e.g., "progress-as-urbanisation", "nature-as-resource") and cultural categories (e.g., "modern/Western vs. traditional/local") embedded in the visuals.

3.5.2. Phase 2: Thematic analysis of student responses

The qualitative questionnaire data were analyzed using reflexive thematic analysis (Braun & Clarke, 2021). The process began with familiarization, during which all responses were read multiple times to gain deep familiarity while recording initial observations. Systematic coding followed, attaching labels to data segments relevant to the research questions; coding remained primarily semantic but also considered latent meanings such as cultural assumptions. NVivo software supported the coding process. After coding, the researcher collated and grouped codes into potential themes, using visual thematic maps to explore relationships. The researcher then reviewed and refined these themes against the coded extracts and the full dataset, merging, splitting, or discarding themes as needed. Once finalized, each theme received a clear definition and a concise, evocative name, along with its underlying "story". Finally, the findings were written up by weaving analytical narrative together with illustrative participant quotes, with Vietnamese quotations translated into English where necessary.

4. Results and discussion

This chapter presents and discusses findings from multimodal discourse analysis of visual artifacts in Prism 3: Listening and Speaking (Lansford et al., 2017) and qualitative survey responses from 37 Vietnamese university students. The chapter has three sections: (1) multimodal analysis of visuals in Unit 4 (Environment) and Unit 5 (Architecture); (2) thematic analysis of student responses (four themes, unchanged from original); (3) integrated discussion.

4.1. Multimodal analysis of visual artifacts

4.1.1. Unit 4: The Environment – Technocratic and Interventionist Schemas

Image 1 (pp. 80-81): Deforestation (A man wearing gloves and noise-cancelling headphones using a chainsaw to cut down a large tree trunk; sawdust is visible flying through the air).

***Table 4.1.1a**. Multimodal analysis of image 1: Deforestation – Man using chainsaw (Prism 3, pp. 80-81)*

Metafunction	Analysis
Representational	A man using a chainsaw to cut a large tree trunk, wearing gloves and noise-cancelling headphones. Flying sawdust emphasizes ongoing destruction. This is a narrative image with a clear action vector – the chainsaw impacts the tree. The man is the actor, the tree is the goal.
Interactive	The man does not look at the camera, creating an "offer" image – the viewer observes the action from a distance. The medium shot captures the full action without intimacy. Protective gear creates an impression of industrial, professional work.
Compositional	The tree trunk occupies the left portion of the frame, the man on the right – creating balance between "human" and "nature", but with humans as the active agent. No forest context; the tree is isolated.

This image opens the unit as a strong symbol of human intervention in nature. From a Cultural Linguistics perspective, it instantiates a cultural schema of the "environment-as-impacted-object" (Sharifian, 2017). The man with modern tools represents human technological power, and the tree represents nature being conquered. However, the image notably lacks any forest context. There is no other trees, no animals, and no community. The tree stands isolated from its ecosystem. This isolation reflects what Merchant (1980) calls the "mechanistic worldview" – a perspective that separates nature into discrete, manageable parts. For Vietnamese learners, forests are not just resources. They are also cultural and spiritual spaces, such as sacred forests or headwater forests (Trần, 2019). Therefore, this framing may feel disconnected from these learners' lived experiences.

***Table 4.1.1b.** Multimodal analysis of the four-image set on cloning endangered species (Prism 3, p. 82)*

Metafunction	Analysis
Image 2a: Banteng calf looking at camera	
Representational	A young banteng calf standing in a barn or research facility, looking directly at the camera. Conceptual, symbolic – the calf represents "hope", "new life", and "clone technology".
Interactive	Direct gaze creates a "demand" image – asking viewers to empathize. Close-up shot creates intimacy.

Metafunction	Analysis
Compositional	Calf occupies almost entire frame, blurred background – complete focus on subject, no environmental context.
Image 2b: Medical team in light blue assisting the calf	
Representational	Four or five people in light blue surgical attire gathered around and assisting the calf. Narrative image with care vectors from humans to animal.
Interactive	No eye contact – viewer observes professional activity. Blue attire and sterile environment suggest scientific practice.
Compositional	Calf at center, people surrounding in a circle – emphasizes focused expertise of medical team.
Image 2c: Scientist with microscope and laboratory equipment	
Representational	Researcher looking into microscope, surrounded by complex lab equipment (test tubes, monitors). Conceptual image symbolizing "research", "high technology".
Interactive	Researcher focused on work, no eye contact – viewer observes from distance. Blue light and dark environment create high-tech, mysterious atmosphere.
Compositional	Complex equipment in foreground, researcher in background – emphasizes technology over human. Shallow depth of field draws attention to microscope and hands.
Image 2d: Reporter with elephants in background	
Representational	Female reporter standing before camera, elephants visible behind in natural setting (grass, trees). Narrative image with communication vector from reporter to audience.
Interactive	Reporter looks directly at camera ("demand") – asking viewers to listen. Elephants in background provide context but not central.
Compositional	Reporter in foreground right, elephants in background left – balance between "media/human" and "animals/nature." Reporter salient due to size and position.

These four images form a clear visual narrative about cloning endangered species. From a Cultural Linguistics perspective, they reflect a schema of "conservation-as-biotechnology" – the idea that the best solution to species extinction lies in scientific laboratories (Sharifian, 2017). This view is deeply rooted in Western traditions that favour technological fixes over social or ecological approaches (Merchant, 1980; Latour, 1993).

The accompanying video adds tension that the images do not resolve. The scientist calls cloning a "miraculous" achievement, while the conservationist argues that ignoring habitat destruction defeats the purpose. The scientist replies that habitat is meaningless without animals to save. Despite this debate, the visuals clearly privilege laboratory settings. Three images show controlled, sterile spaces; only one shows animals in nature, with a reporter in the foreground. This bias reflects what Kopnina (2016) calls "techno-fixing" nature – assuming technology can solve problems that are actually social and political. Escobar (1995) notes that such framings hide alternative knowledges like community-based conservation.

For Vietnamese learners, whose views of nature emphasize reciprocity and spiritual connection (Trần, 2019), this framing feels mismatched. One student said: "*In Vietnam, we talk about the river near my grandmother's house... not about cloning*." The absence of community and traditional knowledge in these visuals marginalizes Vietnamese ecological perspectives.

None of this implies that cloning or laboratory-based conservation should be absent from Academic English materials. These are real and important scientific developments. The issue is one of balance and critical framing: when technocratic solutions appear in the absence of any community-based or traditional alternatives, learners are left with an incomplete picture of how environmental issues are understood and addressed globally.

Table 4.1.1c. *Multimodal analysis of image 3: Waste management workers on coastline (Prism 3, p. 84)*

Metafunction	Analysis
Representational	A group of workers in white protective suits and helmets are collecting and bagging waste on a coastline. Narrative image with action vectors – workers bending, reaching, bagging.
Interactive	Workers focused on task, no eye contact – viewer observes cleanup operation. White suits suggest hazardous materials, professional response.
Compositional	Coastline stretches across frame, waste visible along shore. Workers distributed across scene – coordinated effort. Ocean in background provides context.

This image presents pollution as a technical problem requiring a professional cleanup response. The accompanying text reinforces: "*Conservation of our coastal regions is now vital... This is the result of waste and pollution from the way*

we constantly exploit the habitat." However, the visual solution – workers in hazmat-style suits – is far removed from the everyday reality of waste management in most of the world, including Vietnam. As Bulkeley and Mol (2003) note, environmental governance is increasingly framed as a technical, expert-driven domain, marginalizing community-based and informal practices. For Vietnamese learners, who may experience pollution through daily interactions with plastic waste, informal recyclers, and community cleanups, this image may feel distant and professionalized.

Table 4.1.1d. *Multimodal analysis of image 4: Leopard on Mumbai Streets (Prism 3, p. 87)*

Metafunction	Analysis
Representational	A leopard walking on a street in Mumbai, India, with urban buildings and infrastructure visible. Narrative image of wildlife moving through human-dominated space.
Interactive	Leopard does not look at camera – "offer" image; viewers observe this unusual scene. Nighttime setting and street lighting create a slightly mysterious atmosphere.
Compositional	Leopard occupies center frame, urban background providing context. Contrast between wild animal and built environment is visually striking.

This image, with the text "*Animals are feeling at home in the city and in the countryside*", introduces a schema of urban wildlife adaptation. The choice of Mumbai (non-Western, rapidly urbanizing) is more relatable to Vietnamese learners than purely Western examples. However, as Whatmore (2002) argues, representations of wildlife in cities often still position animals as "out of place" or as intruders into human space, rather than as co-inhabitants. For Vietnamese learners, this image might resonate with local experiences (monkeys in rural areas, birds in cities), but the specific species (leopard) and location maintain some distance from Vietnamese ecological contexts.

Table 4.1.1e. *Multimodal analysis of Image 5: Three Global Climate Change Images – Hurricane, Arctic Melt, Amazon Deforestation (Prism 3, p. 96)*

Metafunction	Analysis
Representational	Satellite hurricane, open Arctic water, cleared Amazon land – iconic global environmental crises.
Interactive	Distant, objective perspectives; invite concern but maintain scientific detachment.
Compositional	Each image fills frame – global scale, no local context or human presence.

These three images present iconic, globally recognized symbols of climate crisis. From a Cultural Linguistics perspective, they instantiate a cultural schema of "climate-change-as-global-crisis" requiring global solutions (Slater, 1995). The satellite perspective of the hurricane creates scientific detachment; the Arctic melt image invites concern through stark loss; the Amazon image documents destruction from an elevated perspective. None show human communities or local perspectives. As Hulme (2009) argues, climate change is often represented through "global imaginaries" that obscure local vulnerabilities and adaptations. For Vietnamese learners, absent images include Mekong Delta saltwater intrusion, urban air pollution, hydropower dam impacts, or coral bleaching – issues directly affecting their communities. This global framing risks making environmental problems feel distant and abstract rather than immediate and personal.

4.1.2. Unit 5: Architecture

The visual artifacts in Unit 5 consistently frame architecture through schemas of progress-as-innovation, urbanisation-as-modernity, and individual achievement. These reflect capitalist ideologies centered on transformation, disruption, and personal success (Harvey, 1989; Berman, 1982). From a Cultural Linguistics perspective, these schemas are not universal but culturally specific, rooted in Western modernist narratives that may not resonate with Vietnamese architectural traditions.

Table 4.1.2a. *Multimodal analysis of Image 6: Modern Architectural Marvel (Prism 3, pp. 102-103)*

Metafunction	Analysis
Representational	A striking image of a modern building with unique, curved architecture, appearing to float or cantilever dramatically. The building stands alone, with minimal contextual surroundings visible. Conceptual image symbolizing "innovation", "modern design", "architectural achievement", and "human ingenuity".
Interactive	The building is presented as an object of admiration and awe. The dramatic angle and clean lines invite aesthetic appreciation. The

	absence of people or everyday activity creates a sense of architecture as art or spectacle rather than lived space.
Compositional	The building dominates the frame, its distinctive shape immediately capturing attention. The sky provides a clean, uncluttered background, isolating the building from any social, cultural, or geographical context.

This opening image establishes architecture as a domain of aesthetic achievement and innovative design. From a Cultural Linguistics perspective, it instantiates a cultural schema of "architecture-as-artistic-statement" – buildings as individual creative expressions rather than as community spaces or cultural containers. The isolation of the building from its surroundings reflects what Kaomea (2003, p. 14) terms "reading erasures" – the systematic omission of contextual factors that would complicate a celebratory narrative of architectural progress. The absence of human figures, streets, vegetation, or neighboring buildings presents architecture as existing in a vacuum, divorced from the social relations and cultural practices that give built spaces meaning. For Vietnamese learners, whose architectural traditions emphasize the integration of buildings with their natural and social environments (Trần, 2019), this framing may appear disconnected from lived experience.

Table 4.1.2b. *Multimodal analysis of image 7 (p. 104): Four Images on "The Skyscraper"*

Metafunction	***Analysis***
Image 7a: 19th-century brownstone (Auditorium Building, Chicago, 1889)	
Representational	*A stately brown stone building representing early skyscraper architecture – the Auditorium Building on Michigan Avenue, designed by Louis Sullivan and completed in 1889. Conceptual image representing "origin", "heritage", "foundation of modern architecture".*
Interactive	*The building is presented as a historical artifact, a precursor to modern skyscrapers. The street-level view emphasizes its presence in the city as a solid, grounded structure.*
Compositional	*The building fills the frame, its classical details and ornate features visible. The angle emphasizes its height relative to surrounding structures, foreshadowing the verticality to come.*
Image 7b: Modern glass-and-steel skyscraper	
Representational	*A contemporary glass-and-steel skyscraper reaching toward the sky, reflective surfaces catching the light. Conceptual image symbolizing "corporate power", "modernity", "economic success", and "technological progress".*
Interactive	*The low-angle (worm's-eye) shot positions the viewer looking up, creating a power differential where the building is accorded authority, dominance, and awe. The reflective glass surface suggests transparency, sophistication, and the erasure of historical ornament.*
Compositional	*Vertical lines draw the eye upward, reinforcing the schema of "upward mobility" and "aspiration". The building dominates the frame, with no contextualizing elements to suggest its relationship to the street, the city, or its inhabitants.*
Image 7c: Middle-aged man in suit looking over city	
Representational	*A middle-aged man in a business suit standing at a window or observation deck, looking down over an expansive cityscape. Conceptual image representing "success", "power", "command", "corporate achievement", and "the male gaze upon the modern city".*
Interactive	*The man's back is turned to the viewer – we see what he sees, invited to share his commanding perspective. The city below represents what he*

Metafunction	*Analysis*
	has achieved or controls, positioning the viewer as a vicarious participant in this power dynamic.
Compositional	*The man occupies the left foreground in sharp focus, his silhouette outlined against the bright city beyond. The city spreads below and beyond him – the human achiever surveying his domain. The contrast between the dark, substantial figure of the man and the vast, sprawling city below reinforces the schema of individual mastery over the urban environment.*
Image 7d: City skyline with river	
Representational	*A wide view of a city skyline with tall buildings of varying heights and a river in the foreground. Conceptual image representing "urban achievement", "economic power", "modern civilization", and "the aesthetic of the skyline as symbol".*
Interactive	*The expansive view invites admiration and perhaps aspiration – this is what a successful city looks like. The river adds aesthetic appeal and suggests commerce, transport, and the historical importance of waterways to urban development.*
Compositional	*The skyline stretches across the frame, with buildings of varying heights creating a distinctive silhouette against the sky. The river in the foreground reflects the buildings, creating a sense of harmony between nature (the river) and human achievement (the skyline) – but only on nature's terms as a decorative backdrop to human progress.*

These four images and the video transcript present the skyscraper as a symbol of American corporate power and global modernity. The transcript states: "*tall buildings represent power*" – an idea that remains popular today. This framing reflects several Western-centric cultural schemas (Sharifian, 2017). First, a "progress-through-technology" schema: the Chicago fire enabled new building techniques, making taller structures possible. Second, an "individual genius" schema: Louis Sullivan as the "father of the skyscraper," ignoring collective processes (Beirne, 2017). Third, a "power-and-success" schema: skylines as "bar charts" of corporate wealth, turning architecture into economic spectacle (Harvey, 1989). Fourth, a "global-expansion" schema: skyscrapers spread from Chicago to Shanghai and Kuala Lumpur, erasing indigenous traditions (King, 2004).

Image 2c – the man in a suit overlooking the city – encodes a schema of "individual mastery." The viewer identifies with his commanding gaze, while the absence of women or community figures reinforces a narrow, masculine view of achievement (Berman, 1982).

For Vietnamese learners, this clashes with the *đình làng* (communal house) – a horizontal, collective, community-focused space (Trần, 2019). As one student said, "*Where is the street? Architecture is not just buildings; it's how people use them.*"

4.2. Thematic analysis of student responses

The qualitative survey responses from 37 Vietnamese university students revealed complex patterns of interpretation, ranging from alignment with the encoded meanings to explicit dissonance and critique.

Four major themes emerged from the reflexive thematic analysis, each revealing important dimensions of how Vietnamese learners engage with Western-centric visual content.

Before presenting these themes, a cautionary note is necessary. The participants in this study are students at an internationally oriented university in Ho Chi Minh City. Their perspectives reflect a particular segment of Vietnamese youth – urban, globally connected, and academically privileged – and should not be read as a monolith for "the Vietnamese learner" or for Vietnamese cultural identity as a whole. The themes below illustrate how this specific group engages with Western-centric visuals, not how all Vietnamese learners would necessarily respond.

4.2.1. Theme 1: Recognition of Western schemas as "global" or "modern"

Many students recognized and, in some cases, accepted the Western cultural schemas embedded in the visuals as representative of "global" or "modern" values. This theme reveals the powerful influence of globalized media and education in shaping Vietnamese youth's cultural conceptualisations.

Student 8 (Female, International Business): "*The skyscraper picture, it looks very modern, very developed. This is what I imagine when I think of a successful country – like Singapore or America. Vietnam is developing, but we don't have many buildings like this yet. It makes me think about the future.*"

Student 23 (Male, Finance): "*The scientist in the lab – this is how they solve environmental problems in developed countries. They have the technology. In Vietnam, we are still struggling with basic pollution. So these images show us what is possible, what we can aim for.*"

Student 15 (Female, Economics): "*The architect looks very professional, very successful. This is the image of a modern career. In Vietnam, we respect teachers and doctors more, but maybe the young generation wants to be like this – creative, independent, successful.*"

These responses reveal students' internalization of a developmentalist discourse where Western technological and architectural achievements are positioned as aspirational targets. The cultural schema of "progress-as-technological-advancement" is recognized and, to varying degrees, accepted. This finding aligns with Vuong et al. (2025), who document how Vietnamese EFL learners often associate "global" with "Western" and "modern". From a Cultural Linguistics perspective, this acceptance suggests that learners are developing distributed cultural cognition that incorporates both local and global schemas – a necessary foundation for metacultural competence (Sharifian, 2013). However, this acceptance also suggests a potential erasure of alternative, locally-rooted definitions of progress. When students automatically equate "modern" with "Western", they may be implicitly devaluing Vietnamese cultural achievements and ways of knowing – a form of cultural cringe that Phillipson (1992) identifies as a consequence of linguistic imperialism.

4.2.2. Theme 2: Experiencing cultural distance and dissonance

A significant number of students expressed feelings of distance, unfamiliarity, or disconnect when engaging with the visuals. This theme reveals the experiential dimension of intercultural mismatch.

Student 14 (Female, International Business): "*The laboratory picture... it feels very far from my life. I have never been in a place like that. In Vietnam, when we talk about environment, we talk about the river near my grandmother's house that is polluted, or the air in Ho Chi Minh City. Not about cloning. It feels like a different world.*"

Student 31 (Male, Economics): "*The architect with the model – he looks very confident, very professional. But I wonder, where is the community? In Vietnam, when we build something, many people are involved – the family, the neighbors, the local government. It's not just one person's vision. This image feels very individualistic, very American.*"

Student 6 (Female, International Business): "*The skyscraper is beautiful, but I don't feel connected to it. In Vietnam, our buildings are lower, closer to the street, closer to people. When I think of architecture, I think of the old apartment buildings where everyone knows each other, not these cold, tall buildings.*"

Student 22 (Male, Economics): "*The endangered animals—they are sad, but they are far away. In Vietnam, we worry about the animals in our own forests, the saola, the langurs. But these pictures show animals I only see in documentaries. It doesn't feel like my problem.*"

These responses reveal intercultural mismatches at the level of cultural schemas. Students' lived experiences – grounded in Vietnamese communal, familial, and local contexts – do not align with the individualist, technocratic schemas encoded in the visuals. This dissonance is not merely affective but cognitive: the learners' distributed cultural cognition cannot easily accommodate the unfamiliar schemas, leading to feelings of disconnection and irrelevance. From a Cultural Linguistics perspective, this is significant because it indicates that the textbook visuals are failing in their communicative function. As Sharifian (2011) argues, effective communication requires alignment between the cultural conceptualisations encoded in a message and those activated in the receiver's cognitive environment. When such alignment fails, pragmatic failure occurs – not in linguistic decoding, but in the interpretation of cultural meaning.

The repeated invocation of Vietnamese specificity – the grandmother's river, the old apartment buildings, the local forests, the saola and langurs – demonstrates that learners are not blank slates but bring rich cultural schemas to the interpretive task. The dissonance they experience arises from the clash between these schemas and those encoded in the textbook. This finding challenges the assumption that global textbooks are culturally neutral or universally applicable, supporting Canagarajah's (1999) critique of linguistic imperialism in ELT materials.

4.2.3. Theme 3: Activating Vietnamese cultural schemas as interpretive resources

When confronted with visuals that felt alien, many students spontaneously activated Vietnamese cultural schemas to make sense of or critique the images. This theme reveals learners' agency in resisting and re-interpreting dominant cultural content.

Student 5 (Female, Economics): "*The endangered animals... they are all separate, alone. In Vietnamese culture, we believe everything is connected. If the forest is destroyed, the animals die, but also the people suffer, the spirits are angry. My grandmother always says, 'Rừng vàng, biển bạc' (golden forest, silver sea) – nature is precious, but it's connected to our lives. These pictures don't show that connection.*"

Student 19 (Male, International Business): "*The skyscraper is beautiful, but where is the street? In Vietnam, life is on the street – the vendors, the motorbikes, the coffee shops.*

Architecture is not just buildings; it's how people use them. This picture shows a building, but it doesn't show life. It feels cold."

Student 28 (Female, International Business): "*The scientist in the lab – in Vietnam, we respect scientists, but we also respect the elders who know about nature. My grandfather never went to university, but he knows which plants can cure sickness, when to plant rice, how to read the weather. That is also knowledge about environment. The picture doesn't show that kind of knowledge.*"

Student 12 (Male, Finance): "*The architect – he is young, individual. In Vietnam, architecture is often about family, about ancestors. When we build a house, we consider feng shui, the directions, the spirits. It's not just one person's idea. It's about harmony with the family and with nature.*"

These responses powerfully demonstrate that Vietnamese learners actively draw on their own cultural schemas – rooted in Confucian, agrarian, and communal traditions – as interpretive resources. The proverb "*Rừng vàng, biển bạc*" (Golden forests and silver seas) cited by Student 5 encapsulates a cultural schema of nature as both precious and integrated with human livelihood, directly challenging the "nature-as-isolated-specimen" schema of the textbook images about cloning. The invocation of street life, grandfather's traditional knowledge, and feng shui principles all instantiate Vietnamese cultural conceptualisations that the textbook visuals ignore.

This finding resonates with Canagarajah's (1999) argument that learners in periphery contexts actively resist and re-interpret dominant cultural content through local frameworks. From a Cultural Linguistics perspective, this is evidence of distributed cultural cognition in action: learners draw on the collective conceptual resources of their cultural community to make meaning. Importantly, this is not mere rejection but a form of critical engagement. Students are not saying the textbook images are "wrong" but that they are incomplete – they tell only one story, ignoring the rich cultural understandings that Vietnamese learners bring to the topics.

This finding has significant pedagogical implications. It suggests that Vietnamese learners possess, and can articulate, the very cultural schemas that the global textbook marginalizes. A culturally responsive pedagogy would build on this resource, using students' cultural knowledge as a bridge to develop metacultural competence. Rather than treating students' cultural frameworks as obstacles to overcome, such a pedagogy would see them as valuable interpretive resources.

4.2.4. Theme 4: Proposing localized visual alternatives

When asked to suggest Vietnamese images that could better represent the unit themes, students offered rich, culturally specific alternatives, demonstrating their capacity for glocalised thinking and their embryonic metacultural competence.

Student 11 (Female, Finance): "*For environment, you could show the rice fields in the Mekong Delta. They are beautiful, but also they show how people live with nature – the water, the plants, the fish, the farmers. It's a system. Or maybe the 'chợ nổi' (floating markets) – they show the river as part of life, not just something to be controlled.*"

Student 27 (Male, Economics): "*For architecture, you should show the 'đình làng' (communal house). In every village, this is the center. It's not modern or tall, but it's where the community meets, where festivals happen, where ancestors are honored. This is Vietnamese architecture. Or maybe the old houses in Hoi An – they mix Vietnamese, Chinese, Japanese styles. That shows our history.*"

Student 33 (Female, International Business): "*I think Hạ Long Bay is perfect for environment and architecture together. The limestone mountains are nature, but the floating villages show how people adapt and live with nature. It's not about controlling nature, it's about finding balance.*"

Student 9 (Male, International Business): "*For environment, maybe show the terraced fields in Sapa. The H'mong people have farmed there for hundreds of years without destroying the mountain. That's sustainable. That's Vietnamese wisdom about nature.*"

Student 18 (Female, Finance): "*For architecture, the Imperial City in Huế shows our history, our kings, our culture. But also the small alleyways in Hanoi – ngõ – where everyone knows each other, where children play, where old people sit and drink tea. That is where Vietnamese life happens, not in tall buildings.*"

These student-generated alternatives are not merely decorative; they embody fundamentally different cultural schemas that merit detailed analysis from a Cultural Linguistics perspective:

Rice fields and floating markets show a schema of "nature-as-life-sustaining-system". In this view, humans participate in nature rather than control it. The Mekong Delta, for example, reflects the concept of *sông nước* (river water) as the source of life, livelihood, and cultural identity. This schema differs clearly from the textbook's "nature-as-resource" or "nature-as-machine". The *đình làng* (communal house) represents "architecture-as-community-container". Here, built space supports social and spiritual togetherness. Unlike the skyscraper's vertical individualism, the *đình làng* is horizontal, open to all, and part of village life. It puts into built form what Trần (2019) calls *tính cộng đồng* (communality). Hạ Long Bay with its floating villages exemplifies *hài hòa với thiên nhiên* (human-nature-harmony). This is a core Vietnamese value rooted in agrarian traditions (Trần, 2019). Unlike the laboratory images that separate humans from

nature, this schema places humans within nature, adapting rather than dominating. Terraced fields in Sapa encode a schema of "sustainable tradition". This is knowledge built over generations, embedded in daily practice, and in balance with natural systems. It directly challenges the textbook's link between environmental solutions and modern scientific expertise. The Imperial City and *ngõ* (urban alleyways) together represent "architecture-as-lived-experience". They include both grand history and everyday life. Unlike the isolated skyscraper, these spaces are relational – they exist through their connection to people, history, and community.

These student examples show that Vietnamese learners have metacultural competence in an early form (Sharifian, 2013). They can recognize, express, and use their own cultural concepts, and compare them with the unfamiliar concepts in the textbook. They can do what Kramsch (2009) calls "third-place" thinking – moving between cultural frameworks instead of being stuck in one. However, this competence remains underdeveloped because the textbook gives no support for such comparative, critical work. Students must deal with cultural mismatch on their own, without clear guidance on how to recognize, analyze, or negotiate between different cultural ways of thinking. This is a missed teaching opportunity and a clear lesson for materials design.

4.3. Integrated discussion: Intercultural mismatches, metacultural competence, and the path forward

4.3.1. The nature of intercultural mismatches

The findings reveal that intercultural mismatches occur at multiple levels of cultural conceptualisation:

At the level of cultural schemas: Vietnamese learners operate with schemas of interconnectedness (between humans and nature, individuals and community) that conflict with the textbook's schemas of isolation and intervention (nature as separate object to be managed, architect as isolated genius). The cloning imagery's schema of "conservation = technological intervention" clashes with students' schema of "conservation = community stewardship". The skyscraper's schema of "progress = vertical achievement" clashes with students' schema of "progress = community wellbeing" and "architecture = lived experience".

At the level of cultural categories: The textbook's categories of "modern architecture" (futuristic, individual, corporate, innovative) and "environmental solutions" (technological, expert-driven, lab-based) exclude the categories that students consider relevant: traditional/vernacular architecture, communal spaces, traditional ecological knowledge, everyday environmental experience, spiritual relationships with nature.

At the level of cultural metaphors: The textbook's implicit metaphors – "NATURE IS A MACHINE TO BE REPAIRED", "CONSERVATION IS LABORATORY WORK", "ARCHITECTURE IS CORPORATE ACHIEVEMENT", "PROGRESS IS UPWARD MOVEMENT" – conflict with Vietnamese cultural metaphors such as "NATURE IS A NETWORK", "CONSERVATION IS COMMUNITY STEWARDSHIP", "ARCHITECTURE IS COMMUNITY CONTAINER", "PROGRESS IS BALANCE".

These mismatches are not superficial; they are foundational to how learners understand the topics under study. When students report that the images feel "cold", "far", "like a different world", they are experiencing a fundamental misalignment between the textbook's cultural conceptualisations and their own distributed cultural cognition.

4.3.2. The paradox of critical voices in textbook content

An important nuance emerges from the analysis of Unit 4. The cloning video includes a critical voice – the conservationist who argues that "*if you don't deal with protecting habitat and dealing with all the root causes of endangerment, it doesn't matter how many animals you are able to produce in the lab*". This could be seen as the textbook including diverse perspectives.

However, from a Cultural Linguistics perspective, the inclusion of this critical voice does not fundamentally alter the dominant framing. The visual imagery overwhelmingly privileges the laboratory, the scientists, the technology. The conservationist's words are present, but the images tell a different story. Moreover, the debate is framed as a technical disagreement between experts, not as an invitation to consider fundamentally different epistemologies – such as indigenous conservation practices, community-based resource management, or spiritual relationships with nature.

This reflects what can be termed "contained diversity" – the inclusion of alternative perspectives within a dominant framework that remains unchallenged. Students hear that habitat protection matters, but they see laboratories and cloning technology. The underlying schema – that solutions come from experts, that technology is central – remains intact.

4.3.3. Implications for metacultural competence and pragmatic failure

To specify the cognitive mechanism at play, this study identifies two interrelated processes. First, learners experience pragmatic failure at the conceptual level (Thomas, 1983): they decode the literal content of the image (e.g., "this is a cloning laboratory") but fail to align with its intended cultural framing (e.g., "conservation is primarily a technological enterprise"). This misalignment occurs because

the textbook assumes a Western "nature-as-machine" schema that is not part of the learners' distributed cultural cognition. Second, learners simultaneously experience identity friction (Sharifian, 2017): they understand the Western schema perfectly yet feel a sense of dissonance because the visual bypasses their indigenous Confucian and agrarian schemas. As one student noted, the cloning lab "feels like a different world" – not because she misunderstood it, but because it did not resonate with her own cultural framework. These two mechanisms – pragmatic failure and identity friction – operate together, making the visuals cognitively accessible but culturally alienating.

From a Cultural Linguistics perspective, these mismatches have direct implications for learners' communicative competence. Sharifian (2013) argues that effective communication in English as an international language requires metacultural competence – the ability to consciously recognize, interpret, negotiate, and employ different cultural conceptual systems. The findings suggest that Vietnamese learners are currently developing this competence in an ad hoc, unsupported manner. They recognize mismatches, activate local schemas, and propose alternatives, but they do so without explicit guidance or structured opportunities to develop these skills.

This has consequences for pragmatic competence in cross-cultural communication. When Vietnamese learners encounter Western interlocutors who operate with schemas of "progress-as-innovation" or "nature-as-resource", they may misinterpret communicative intent or struggle to express their own culturally grounded perspectives. For example, a Vietnamese professional discussing urban development with Western colleagues might intuitively frame the discussion around community needs and social harmony, while the Western colleagues operate with schemas of individual achievement and technological innovation. Without metacultural awareness – the ability to recognize these different schemas and consciously negotiate between them – miscommunication and pragmatic failure are likely. The textbook's failure to address these cultural dimensions of meaning thus has real-world consequences for learners' communicative effectiveness.

4.3.4. Learner agency as a resource, not a deficit

The qualitative data reveal an important nuance that challenges any victim narrative. Students in this study were not passive recipients of Western content who simply felt "alienated". Instead, they actively negotiated meaning by invoking their own cultural schemas – proverbs, family knowledge, street life, spiritual beliefs. In doing so, they demonstrated that the very mismatch between textbook visuals and local schemas can stimulate metacultural competence rather than merely hinder it. For example, when Student 5 cited the proverb "Rừng vàng, biển bạc" (Golden forests and silver seas) to critique the isolated endangered species images, she was performing exactly the kind of cross-cultural comparison that Sharifian (2013) describes as core to metacultural competence. The friction between the textbook's framing and her own cultural knowledge did not paralyze her; it prompted critical reflection and articulation. Therefore, while the textbook provides no scaffolding for such work, the learners themselves possess the raw materials for developing metacultural awareness. Pedagogical intervention should aim to harness this existing capacity, not to remedy a deficit.

5. Conclusion

5.1. Summary of key findings

This study investigated the role of visual artifacts in the academic English textbook Prism 3: Listening and Speaking (Lansford et al., 2017), a widely adopted resource in Vietnamese tertiary EFL contexts. Employing the framework of Cultural Linguistics (Sharifian, 2011, 2017) and Multimodal Discourse Analysis (Kress & van Leeuwen, 2006), the study analyzed 14 visual artifacts from Unit 4 (The Environment) and Unit 5 (Architecture), supplemented by qualitative survey responses from 37 Vietnamese university students. The key findings are summarized below:

First, Western-centric cultural conceptualisations are systematically encoded in the visual artifacts of Prism 3. Through representational, interactive, and compositional choices, the visuals privilege specific cultural schemas, categories, and metaphors. In Unit 4, environmental issues are framed through a "technocratic-environmental schema" that positions nature as an object to be managed, controlled, or repaired through scientific expertise and technological intervention. Representations of deforestation, cloning laboratories, waste management, and global climate change imagery all reflect this schema. In Unit 5, architectural issues are framed through a "progress-as-innovation schema" that privileges individual achievement, corporate success, technological novelty, and the global spread of Western architectural forms – embodied most clearly in the skyscraper as a symbol of economic power.

Second, these encoded schemas conflict with Vietnamese learners' cultural conceptualisations rooted in Confucian collectivism, agrarian ecological balance, and communitarian values. The multimodal analysis identified consistent patterns of marginalization: representations of communities living in reciprocal relationship with nature, traditional ecological knowledge, architecture as lived everyday experience, and spiritual or cultural dimensions of both environment and architecture are systematically absent from the textbook's

visual discourse.

Third, Vietnamese learners experience cultural dissonance when engaging with these visuals. Survey responses revealed feelings of distance, unfamiliarity, and disconnection. Students explicitly articulated mismatches between their lived experiences – polluted rivers, local forests, street life, communal spaces, traditional knowledge – and the unfamiliar schemas presented in the textbook. This dissonance is not merely affective but cognitive; it signals a failure of alignment between the cultural conceptualisations encoded in the visuals and those activated in learners' distributed cultural cognition (Sharifian, 2011).

Fourth, learners actively activate Vietnamese cultural schemas as interpretive resources. Far from being passive recipients, students demonstrated agency in resisting and re-interpreting dominant content. They invoked proverbs (e.g., "Rừng vàng, biển bạc" - Golden forests and silver seas), traditional practices (feng shui, elder knowledge), lived experiences (street life, grandmother's river), and spiritual connections to nature. This critical engagement challenges deficit models of learners in periphery contexts.

Fifth, students articulated rich, localized visual alternatives. Proposals included rice fields and floating markets in the Mekong Delta, the *đình làng* (communal house), Hạ Long Bay with floating villages, terraced fields in Sapa, the Imperial City in Huế, and *ngõ* (urban alleyways) in Hanoi. These alternatives embody fundamentally different cultural schemas – "nature-as-life-sustaining-system, architecture-as-community-container, human-nature harmony, sustainable tradition, and architecture-as-lived-experience" – that directly challenge the textbook's dominant framings.

Sixth, current materials fail to develop metacultural competence. While students demonstrated embryonic metacultural awareness – recognizing mismatches, activating local schemas, proposing alternatives – the textbook provides no scaffolding for explicitly recognizing, analyzing, or negotiating between different cultural conceptual systems. Students navigate cultural dissonance unsupported, representing a missed pedagogical opportunity.

5.2. Theoretical Implications

This study makes four contributions. First, it shows the value of combining Cultural Linguistics with Multimodal Discourse Analysis. Visual artifacts carry cultural schemas as powerfully as language does, and the integration of MDA provides a systematic way to analyze how cultural meanings are built through representational, interactive, and compositional choices. Second, it extends Cultural Linguistics to ELT materials analysis, revealing how global textbooks privilege certain worldviews. Third, it contributes to critical applied linguistics by showing how linguistic imperialism (Phillipson, 1992) operates through visual channels, and introduces the concept of "contained diversity" – critical voices included but not challenging the dominant framework. Fourth, it refines understanding of metacultural competence (Sharifian, 2013): learners have it in early form but need explicit scaffolding to develop full awareness.

5.3. Pedagogical implications

5.3.1. Materials design

The study supports glocalised approaches (Robertson, 1995) but argues that supplementation alone is not enough. A shift toward conceptual pluralism is needed. Designers should include non-Western architectural traditions and local environmental perspectives (rice fields, floating markets) alongside global ones, create tasks that compare different cultural conceptualisations, and avoid "contained diversity" by representing alternatives visually.

5.3.2. Classroom practice

Teachers can adopt a metacultural pedagogy using strategies such as consciousness-raising (asking whose values images assume), schema elicitation (how would students represent the topic from a Vietnamese perspective?), comparative analysis, negotiation practice, and critical response tasks.

Teachers can turn cultural mismatches into active learning opportunities using five strategies suggested by student responses. First, juxtaposition and comparison: present a textbook image alongside a local alternative (e.g., a skyscraper next to a *đình làng*) and ask students to compare the values each conveys. Second, critical questioning: use guided prompts (e.g., "How do your family or elders protect nature?") to help students articulate their own cultural schemas. Third, localization tasks: ask students to find or create Vietnamese images (rice fields, floating markets, Hạ Long Bay) to supplement or replace textbook visuals, then use them for presentations or writing assignments. Fourth, role-play and negotiation: simulate cross-cultural meetings where students explain a local design (e.g., a low-rise communal house) to an international client expecting a glass tower. Fifth, reflexive writing: have students keep a "cultural friction journal" to record moments of mismatch and propose alternative images or concepts.

5.3.3. Teacher education

Vietnamese EFL teachers need training in recognizing cultural conceptualisations in textbooks, developing glocalised supplements, facilitating metacultural discussions, and critically evaluating materials.

5.4. Limitations

First, the study analyzed only one textbook and two units. Second, it involved only 37 students from one university (These students from a single urban university possess high levels of globalized cultural capital, which means their critical awareness of Western-centric visuals may be more developed than that of learners in other Vietnamese contexts). Third, survey responses may not capture immediate interpretations. Fourth, the researcher's insider status may have introduced bias. Fifth, the cross-sectional design does not show change over time.

5.5. Directions for future research

Future research could pursue six directions: comparative analysis of multiple global and local textbooks; classroom-based intervention studies; in-depth qualitative methods (interviews, focus groups); research on teacher perspectives; extension to other modalities (layout, audio, video); and cross-cultural replication in other non-Western EFL contexts.

References

[1] Beirne, M., Jennings, M., & Knight, S. (2017). Autonomy and resilience in cultural work: Looking beyond the 'creative industries'. *Journal for Cultural Research, 21*(2), 204-221. https://doi.org/10.1080/14797585.2016.1275311

[2] Berman, M. (1982). *All that is solid melts into air: The experience of modernity*. Simon and Schuster.

[3] Bourdieu, P. (1991). *Language and symbolic power* (G. Raymond & M. Adamson, Trans.). Harvard University Press.

[4] Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage Publications.

[5] Bulkeley, H., & Mol, A. P. J. (2003). Participation and environmental governance: Consensus, ambivalence and debate. *Environmental Values, 12*(2), 143-154. https://doi.org/10.3197/096327103129341270

[6] Canagarajah, A. S. (1999). *Resisting linguistic imperialism in English teaching*. Oxford University Press.

[7] Chen, Y. (2010). The semiotic construal of attitudinal curriculum goals: Evidence from EFL textbooks in China. *Linguistics and Education, 21*(1), 60–74. https://doi.org/10.1016/j.linged.2009.12.005

[8] Crookes, G. V. (2013). *Critical ELT in action: Foundations, promises, praxis*. Routledge.

[9] Denzin, N. K., & Lincoln, Y. S. (Eds.). (2018). *The Sage handbook of qualitative research* (5th ed.). Sage Publications.

[10] Flick, U. (2018). *An introduction to qualitative research* (6th ed.). Sage Publications.

[11] Harvey, D. (1989). *The condition of postmodernity*. Blackwell.

[12] Holliday, A. (2006). Native-speakerism. *ELT Journal*, *60*(4), 385-387. https://doi.org/10.1093/elt/ccl030

[13] Hulme, M. (2009). *Why we disagree about climate change: Understanding controversy, inaction and opportunity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511841200

[14] Jewitt, C. (Ed.). (2009). *The Routledge handbook of multimodal analysis*. Routledge.

[15] Kaomea, J. (2003). Reading erasures and making the familiar strange: Defamiliarizing methods for research in formerly colonized and historically oppressed communities. *Educational Researcher*, *32*(2), 14-25. https://doi.org/10.3102/0013189X032002014

[16] King, A. D. (2004). *Spaces of global cultures: Architecture, urbanism, identity*. Routledge.

[17] Kramsch, C. (2009). *The multilingual subject*. Oxford University Press.

[18] Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

[19] Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold Publishers.

[20] Kress, G., & van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.

[21] Lansford, L., Lockwood, R. B., Cavage, C., & Blackwell, A. (2017). *Prism 3: Listening and speaking*. Cambridge University Press.

[22] Merchant, C. (1980). *The death of nature: Women, ecology, and the scientific revolution*. Harper & Row.

[23] Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation*(4th ed.). Jossey-Bass.

[24] Pennycook, A. (2017). *The cultural politics of English as an international language*. Routledge. (Original work published 1994)

[25] Phillipson, R. (1992). *Linguistic imperialism*. Oxford University Press.

[26] Robertson, R. (1995). Glocalization: Time-space and homogeneity-heterogeneity. In M. Featherstone, S. Lash, & R. Robertson (Eds.), *Global modernities* (pp. 25-44). Sage Publications.

[27] Rose, G. (2016). *Visual methodologies: An introduction to researching with visual materials* (4th ed.). Sage Publications.

[28] Serafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. Teachers College Press.

[29] Setyono, B., & Widodo, H. P. (2019). The representation of

multicultural values in the Indonesian Ministry of Education and Culture-endorsed EFL textbook: A critical discourse analysis. *Intercultural Education*, *30*(4), 383-397. https://doi.org/10.1080/14675986.2019.1548102

[30] Sharifian, F. (2011). *Cultural conceptualisations and language: Theoretical framework and applications*. John Benjamins Publishing Company. https://doi.org/10.1075/clscc.1

[31] Sharifian, F. (2013). Globalisation and developing metacultural competence in learning English as an International Language. *Multilingual Education*, *3*(7), 1-11. https://doi.org/10.1186/2191-5059-3-7

[32] Sharifian, F. (2017). *Cultural linguistics: Cultural conceptualisations and language*. John Benjamins Publishing Company. https://doi.org/10.1075/clscc.8

[33] Slater, D. (1995). Photography and modern vision: The spectacle of 'natural magic'. In C. Jenks (Ed.), *Visual culture* (pp. 218-237). Routledge.

[34] Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443-466). Sage Publications.

[35] Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, *4*(2), 91-112. https://doi.org/10.1093/applin/4.2.91

[36] Trần, N. T. (2019). *Tìm về bản sắc văn hóa Việt Nam: Cái nhìn hệ thống - loại hình* (11th ed.). Ho Chi Minh City General Publishing House.

[37] Vuong, K. T. M., Takulasemsuk, W., & Zahabi, A. (2025). English textbooks in the modern world: Cultural representation in an EFL textbook from Vietnam. *PASAA*, *70*(1), 204-238. https://doi.org/10.58837/CHULA.PASAA.70.7

[38] Whatmore, S. (2002). *Hybrid geographies: Natures cultures spaces*. Sage Publications.