

Visual Narratives in EFL Textbooks: A Multimodal Approach to Enhancing Cross-Cultural Awareness and Second Language Acquisition

Waner Bin

School of International Studies, Hangzhou Normal University, No. 2318 Yuhangtang Road, Yuhang District, Hangzhou, China

Email: binwaner0116@163.com

How to cite this paper: Bin, W. N. (2025). Visual Narratives in EFL Textbooks: A Multimodal Approach to Enhancing Cross-Cultural Awareness and Second Language Acquisition. *International Journal of Social Science, Education and Humanities*, 1(1), 41-49. ISSN Print: 3104-4239; ISSN Online: 3104-4247. <https://doi.org/10.63313/IJSSEH.9005>

Published: 2025-08-11

Copyright © 2025 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Abstract

In an era of globalization, effective language education must transcend mere linguistic proficiency to cultivate deep cross-cultural awareness. This study proposes the integration of visual narratives into English as a Foreign Language (EFL) textbooks through a multimodal lens, combining Multimodal Discourse Analysis (MDA) with Sociocultural Theory (SCT) and Vygotsky's Zone of Proximal Development (ZPD). By analyzing the use of composition, color, and character interactions within visual narratives, the research reveals that traditional EFL textbooks often provide superficial cultural representations, whereas dynamic visual narratives offer a more immersive and cognitively engaging learning experience. The findings indicate that scaffolded instructional strategies leveraging these multimodal resources can significantly enhance both cultural understanding and second language acquisition, suggesting a promising framework for future curriculum design in EFL contexts.

Keywords

visual narratives; multimodal discourse analysis; cross-cultural awareness; second language acquisition; EFL textbooks; sociocultural theory

1. Introduction

In the context of contemporary globalization, the evolving intercultural landscape demands diversity and systematic development of cross-cultural competencies within language education (Kramsch, 1993). Contemporary language learners must therefore cultivate not only linguistic proficiency but also the cognitive flexibility to navigate multiple cultural frameworks (Byram, 1997). Current scholarship positions cross-cultural awareness as a pedagogical cornerstone rather than supplementary component, recognizing its critical role in facilitating effective intercultural communication within globalized contexts (Deardorff, 2006).

EFL (English as a Foreign Language) textbooks, as the core medium of language instruction, undertake the dual responsibility of fostering language proficiency and shaping cultural awareness. However, current EFL textbooks still exhibit several notable limitations. One point is the superficial cultural representation, which means that many textbooks present cultural elements through isolated text fragments or standalone images. This often fails to capture the dynamic and complex nature of culture, leading to a superficial understanding of the target culture among learners (Zhang & Yu, 2020). The other is the underutilization of multimodal resources. Although modern pedagogical theories emphasize the integration of various modes of communication, the potential of visual narratives as a composite medium integrating images, symbols, and text remains underdeveloped in EFL textbooks (Kress & van Leeuwen, 2006). This limitation restricts the effectiveness of textbooks in guiding learners to explore cultural nuances.

“Visual narratives” in multimodal discourse analysis is based on visual grammar (Kress & van Leeuwen, 2006), which refers to image or sequences of images that tell stories, as seen in picture books in narrative form with various elements like characters and actions (Painter, et al., 2013). In fostering cross-cultural awareness, visual narratives exhibit unique advantages. Firstly, aligned with Vygotsky’s Zone of Proximal Development (1978), visual storytelling provides structured learning opportunities. Tasks like analyzing cultural symbols or predicting story outcomes guide learners from basic awareness to critical thinking (Lantolf & Thorne, 2006).

In contrast, pure text or static images have limitations in conveying cultural connotations. Visual storytelling integrates images, context, and narrative, activating visual, emotional, and cognitive processing. This immersive approach boosts retention and fosters empathy, helping learners connect more deeply with cultural content (Norris & Ortega, 2000). Moreover, research indicates that visual narratives play a significant role in cross-cultural communication. For instance, in the international dissemination of Chinese stories, optimizing narrative content and dissemination strategies can effectively enhance cross-cultural awareness and understanding (Yang & Du, 2024). Learners engage in activities like interpreting nonverbal cues or role-playing cultural misunderstandings, enhancing their readiness for real-world communication.

In summary, traditional texts rely on a single-channel linguistic input, making them susceptible to learners’ language proficiency limitations. Static images, on the other hand, lack narrative coherence, making it difficult to convey cultural dynamism (Zhang & Yu, 2020). In contrast, visual narratives leverage the synergy between image sequences and text (Painter et al., 2013) to simulate authentic cross-cultural scenarios and activate multimodal cognitive processing (Norris & Ortega, 2000), offering irreplaceable advantages. Through their unique presentation methods and multimodal characteristics, they can more effectively cultivate students’ cross-cultural awareness, surpassing the limitations of traditional text or static im-

ages.

While cross-cultural competence is widely regarded as essential and multimodal teaching methods have demonstrated considerable potential, there remains a significant lack of research on visual narratives in EFL textbooks, particularly in relation to cross-cultural awareness and second language acquisition. Existing research on multimodality has predominantly focused on the analysis of static images, while studies grounded in sociocultural theory have primarily explored classroom interactions and teacher guidance. However, recent empirical findings suggest that a more targeted examination of visual materials in EFL classrooms can significantly enhance students' understanding and appreciation of the target culture (Kiss & Weninger, 2017).

Although the partial validation of relevant theoretical frameworks, large-scale empirical research on textbook design that integrates dynamic visual narratives remains limited. Nonetheless, recent studies, such as those by Navila, Rochsantiningih, and Drahati (2023), highlight the feasibility and positive impact of applying digital multimodal composing frameworks in the development of digital storytelling books. These findings underscore the potential for incorporating similar approaches into EFL education to enrich cross-cultural learning experiences.

In sum, on the one hand, multimodal research has long focused on static image analysis (Kress & van Leeuwen, 2006), whereas dynamic visual narratives require the integration of narratology, visual grammar, and technological tools, posing greater demands for interdisciplinary collaboration. Therefore, despite its recognized potential, its practical application remains lagging behind. On the other hand, with the widespread adoption of digital tools, the development costs of dynamic visual narratives have decreased, providing practical feasibility for this study. Based on the above issues, this study aims to address the following core research questions:

- 1) How do visual narratives in EFL textbooks utilize multimodal resources such as composition, color, and character interactions to promote learners' cross-cultural awareness ?
- 2) How can visual narratives, based on sociocultural theory, enhance second language acquisition efficacy through scaffolded interactions ?
- 3) How can visual narrative design be optimized according to the Zone of Proximal Development (ZPD) theory to accommodate learners' varying levels of cultural cognition ?

The innovation of this study lies in integrating multimodal discourse analysis with sociocultural theory to construct a comprehensive model for EFL textbook design. This model serves three primary functions. First, it examines how dynamic visual narratives in EFL textbooks employ multimodal resources to convey rich cultural information. Second, it explores how scaffolded interactions, grounded in sociocultural theory, can support students in internalizing cultural knowledge while en-

hancing their language proficiency. Finally, it proposes a framework for designing graded visual narrative tasks based on the Zone of Proximal Development (ZPD) theory, enabling a progressive advancement in learners' cross-cultural awareness and second language acquisition.

2. Establishing the Foundation: Multimodal Discourse Analysis and Visual Grammar

In recent years, there has been a fundamental shift in the understanding of communication, recognizing it as inherently multimodal. Multimodal Discourse Analysis (MDA), as articulated by Kress and van Leeuwen (2006), posits that meaning construction relies not only on text but also on the interaction of various semiotic resources such as images, symbols, and layout design. This theoretical framework provides a powerful tool for analyzing how visual elements convey meaning. Visual grammar, a core component of MDA, deconstructs images into three primary dimensions:

Representational Meaning: Refers to how images reflect cultural practices by depicting characters, actions, and scenes. For example, the spatial arrangement of certain images in textbooks may evoke collectivist values (Kress & van Leeuwen, 2006).

Interactive Meaning: Research by Painter et al. (2013) indicates that the perspective and color saturation in images can influence viewers' emotional responses and engagement levels. A horizontal perspective may imply equality, while vivid colors can evoke emotional resonance.

Compositional Meaning: The arrangement of text and images guides viewers' attention, forming a narrative flow. Effective compositional strategies, such as contrasting image arrangements, can prompt learners to identify and critically evaluate cultural differences.

Current research on multimodal discourse analysis primarily focuses on the definition, development, and hotspots of multimodal discourse. However, these studies exhibit certain limitations in presenting cultural complexity. For instance, many studies share similar subjects, unevenly distributed research topics, limited research levels, and a lack of interdisciplinary integration (Zhang et al., 2023).

Visual narratives, as an essential component of multimodal discourse analysis, can intuitively and vividly convey cultural information through visual elements such as images and videos, compensating for the shortcomings of traditional text analysis in presenting cultural complexity. By analyzing the representational, interactive, and compositional meanings of visual symbols, researchers can gain a deeper understanding of the meanings constructed by images from the perspective of visual grammar and the intentions of the creators (Yang & Zhou, 2023). Furthermore, current empirical research highlights the need for improvements in textbook design. For example, some studies point out that the cultural content in English textbooks primarily focuses on the target language culture, with limited representation of oth-

er cultures and local cultures, and a lack of diversity in cultural presentation forms (Zhang & Yu, 2020). These findings directly indicate the need for improvements in the selection and presentation of cultural content in textbook design to more comprehensively reflect cultural diversity and complexity.

3. The Close Connection Between Multimodality and Sociocultural Theory

Multimodal discourse analysis provides tools for analyzing the ideographic functions of visual symbols, while SCT guides how to promote cultural internalization through scaffolded interaction. The combination of the two enables this study to not only deconstruct the multimodal resources in textbooks but also design teaching tasks that are in line with the ZPD. As can be seen from the structure:

[MDA Analysis] → [SCT-Guided Teaching] → [ZPD-Based Task Design]

Drawing on the fundamental ideas of Vygotskian theory (1978), which have been further developed in sociocultural research (Lantolf, 2000), socio-cultural theory (SCT) places significant emphasis on the crucial role of mediated learning and the Zone of Proximal Development (ZPD). These elements support learners in making the transition from dependence to autonomy via scaffolded instruction. Within this framework, digital multimodal resources, particularly visual narratives, can serve dual roles: they function as vehicles for transmitting diverse cultural symbols and practices while also acting as cognitive tools that foster critical analysis and deeper meaning-making (O'Halloran, 2008).

Empirical research supports these theoretical insights. For example, in the studies by Kalantzis and Cope (2009), the multiliteracies perspective demonstrates that a methodical examination of the relationships between images and texts can enhance semiotic awareness and foster a deeper cultural understanding. Moreover, research on digital communication and Web 2.0 practices indicates that scaffolded multimodal instruction creates an instructional continuum, guiding learners from surface-level symbol recognition toward comprehensive cultural internalization (Warschauer & Grimes, 2007).

Despite these promising developments, the application of multimodal principles in EFL textbook design remains inconsistent. Although multimodal resources are acknowledged for their potential in cultural education, their implementation is often fragmented and under-theorized. In fact, findings suggest that many language textbooks continue to offer superficial cultural representations, with isolated images and decontextualized texts failing to capture the dynamic nature of cultural interaction (Deardorff, 2006). These challenges underscore the need for innovative textbook designs that leverage the coordinated use of multiple modes to enhance cultural meaning construction and promote learners' critical engagement with cultural content.

4. From Theory to Practice: The Rationality of Current Research

Understanding how theoretical frameworks translate into effective pedagogical strategies is essential for enhancing both linguistic competence and intercultural awareness in EFL classrooms. While previous chapters have outlined the theoretical underpinnings of multimodal discourse analysis and sociocultural theory, this chapter focuses on the practical application of these theories in instructional design. Specifically, it examines how scaffolded teaching and multimodal analysis work together to support students in interpreting cultural connotations embedded in visual narratives. Through this integration, educators can design learning experiences that not only develop students' language skills but also foster their ability to engage critically with cultural content. The following sections explore how scaffolded instructional approaches and multimodal analysis can be applied synergistically to achieve these goals.

4.1. The Role of Scaffolded Instructional Design in Exploring Cultural Connotations

Scaffolded instructional design is central to enabling EFL learners to negotiate complex cultural connotations embedded in visual narratives. Effective scaffolding through pre-teaching, guided analysis, and interactive discussion can help students move from surface-level recognition of images to a deeper understanding of the cultural and ideological nuances behind them. Recent studies have underscored that when teachers provide targeted support, learners are better able to critically analyze symbolic elements (e.g., colors, motifs, spatial arrangements) and connect these with broader sociocultural contexts. For instance, Chen (2020) discovers that the roles of digital scaffolding materials in language learning and cultural knowledge learning are different. Such scaffolded approaches thus serve both as cognitive bridges and as means to empower learners to question and reinterpret cultural messages in the textbooks.

4.2. The Synergistic Effect of Multimodal Analysis and Scaffolded Teaching

The combination of multimodal analysis and scaffolded teaching can construct a complete teaching chain from symbol interpretation to cultural internalization. Multimodal analysis provides tools for identifying and interpreting cultural symbols, while scaffolded teaching offers students the support needed to gradually internalize cultural meanings during the interpretation process.

In EFL classrooms, teachers can start with a multimodal text, like a storybook with pictures. Ouyang et al. (2022) applies multimodal learning analytics to study instructor scaffoldings' effects on small groups' collaborative programming in K-12, finding various scaffoldings and responses, highlighting the importance of multimodal analysis and scaffolded teaching combination. Multimodal analysis helps students notice details such as character expressions and colors, which might rep-

resent cultural symbols. Then, through scaffolded tasks, teachers guide students. The integration of dynamic visual narratives in EFL textbooks with scaffolded interactions and graded tasks forms a comprehensive framework that enhances both language acquisition and cross-cultural awareness. The integration of theories proposed in this study establishes a powerful pedagogical cycle for handling visual narratives:

1) **Multimodal Deconstruction:** By applying Multimodal Discourse Analysis (MDA) frameworks, teachers and learners can methodically break down and analyze the connections between visual and textual elements. This helps in understanding how different semiotic resources, such as images, symbols, and text, interact to convey meaning.

2) **Cultural Semiosis:** With the help of Sociocultural Theory (SCT), teachers can provide structured guidance to students. This enables students to interpret the symbolic meanings within visual narratives. Scaffolded instruction based on SCT supports learners in moving from basic understanding to a more in - depth exploration of cultural symbols.

3) **Intercultural Internalization:** To promote cross-cultural understanding, reflective tasks are incorporated. These tasks encourage learners to take on different cultural perspectives. By doing so, students can internalize cultural knowledge, enhancing their cross-cultural awareness and making the learning experience more meaningful.

In short, this approach makes lessons more engaging and effective. It not only helps students learn new words and phrases but also lets them explore and understand different cultures, all in a supportive and gradually challenging environment. By integrating multimodal analysis with scaffolded teaching in sociocultural theory, teaching activities can more effectively guide students from surface-level symbol interpretation to deep cultural connotation exploration and internalization, promoting students' cross-cultural communication skills.

5. Conclusion

This study explores how incorporating visual narratives into EFL textbooks can make language learning more engaging and culturally meaningful. By combining multimodal discourse analysis (MDA) with sociocultural theory (SCT), it addresses key shortcomings in traditional textbooks, such as their often shallow cultural content and limited use of multimodal resources. The central argument is that visual narratives, which blend images, symbols, and text, can create richer learning experiences, helping students develop both language skills and cross-cultural awareness. The study draws on Kress and van Leeuwen's (2006) visual grammar and Vygotsky's (1978) theory of scaffolded learning. By breaking down visual narratives into representational, interactive, and compositional elements, teachers can help students interpret cultural symbols and develop critical thinking skills. At the same

time, SCT' s focus on guided learning and the Zone of Proximal Development (ZPD) offers a structured way to gradually introduce more complex cultural concepts, making them easier for learners to absorb. This dual approach not only improves language proficiency but also helps students think more flexibly about cultural differences.

In today's globalized world, there is a growing need for language education to reflect the fluid and interconnected nature of cultures. Traditional EFL textbooks often rely on static images and isolated cultural references, which fail to capture real-life cultural dynamics. Visual narratives, on the other hand, allow learners to engage with cultural scenarios in a more immersive way. Activities like analyzing body language, predicting story outcomes, or role-playing cultural misunderstandings align with research showing that scaffolded, task-based learning is highly effective (Norris & Ortega, 2000; Kiss & Weninger, 2017).

What makes this study stand out is its interdisciplinary approach, combining MDA and SCT to create a more holistic model for EFL textbook design. By tailoring visual narrative tasks to different proficiency levels, the framework ensures that students with varying levels of cultural knowledge can engage meaningfully. Additionally, the study considers the role of digital tools, building on recent research on digital storytelling (Navila et al., 2023), to show how modern technology can make EFL materials more relevant and engaging.

Ultimately, this research calls for a shift in EFL teaching, encouraging educators and curriculum designers to integrate visual narratives into language instruction. This approach has the potential to make language learning more immersive, culturally rich, and adaptable to real-world communication. Future studies will be essential to fine-tune the framework and explore how it can be applied effectively across different educational settings.

References

- [1] Byram, M. (1997). Teaching and assessing intercultural communicative competence. *Multilingual Matters*.
- [2] Chen, C. (2020). Using digital scaffolding materials to enhance learners' autonomous learning: A case in a Chinese university. *Academic Journal of Humanities & Social Sciences*, 3(8), 82 – 96.
- [3] Deardorff, D. K. (2006). The identification and assessment of intercultural competence as a student outcome of internationalization at institutions of higher education in the United States. *Journal of Studies in International Education*, 10(3), 241 – 266.
- [4] Kalantzis, M., & Cope, B. (2009). Multiliteracies, new literacies, and digital epistemologies. *Journal of Literacy Research*, 41(4), 403 – 424.
- [5] Keles, U., & Yazan, B. (2023). Representation of cultures and communities in a global ELT textbook: A diachronic content analysis. *Language Teaching Research*, 27(5), 1325-1346.
- [6] Kiss, T., & Weninger, C. (2017). Cultural learning in the EFL classroom: The role of visuals. *ELT Journal*, 71(2), 186 – 196.
- [7] Kramsch, C. (1993). *Context and culture in language teaching*. Oxford University Press.

- [8] Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design* (2nd ed.). Routledge.
- [9] Lantolf, J. P. (2000). Sociocultural theory and second language learning. *Annual Review of Applied Linguistics*, 20, 191 – 214.
- [10] Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the genesis of second language development*. Oxford University Press.
- [11] Liu, A. H. (2017). A study on the cultivation of intercultural communicative competence in college English under the multimodal interactive teaching model. *Journal of Chongqing University of Arts and Sciences (Social Sciences Edition)*, (03), 88 – 91.
- [12] Navila, A., Rochsantiningasih, D., & Drajati, N. A. (2023). EFL pre-service teachers' experiences using a digital multimodal composing framework to design digital storytelling books. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 8(2), 217 – 231.
- [13] Norris, J., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417 – 528.
- [14] O'Halloran, K. (2008). Multimodal discourse analysis in the classroom: A case study approach. *Journal of Pragmatics*, 40(12), 1932 – 1952.
- [15] Ouyang, F., Dai, X., & Chen, S. (2022). Applying multimodal learning analytics to examine the immediate and delayed effects of instructor scaffoldings on small groups' collaborative programming. *International Journal of STEM Education*, 9(1), 45.
- [16] Painter, C., Martin, J., & Unsworth, L. (2013). *Reading visual narratives: Image analysis of children's picture books*. Equinox.
- [17] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- [18] Warschauer, M., & Grimes, D. (2007). Audience, authorship, and artifact: The emergent semiotic labor of Web 2.0. *Language Learning & Technology*, 11(3), 1 – 23.
- [19] Yang, X. C., & Zhou, Z. X. (2023). Multimodal discourse analysis based on visual grammar. *Journalism and Communication Sciences*, (3), 416 – 423.
- [20] Yang, X. M., & Du, X. (2024). A narrative study of telling China's stories from an intercultural perspective. *Journal of Qingdao University of Science and Technology (Social Sciences Edition)*, 40(1), 70 – 75.
- [21] Zhang, D., Ren, J., Li, X. A., & Wang, W. Y. (2023). A review of multimodal discourse studies. *Modern Linguistics*, 1, 11(11), 5327 – 5332.
- [22] Zhang, H., & Yu, R. (2020). A study on the representation of Chinese culture in college English textbooks. *Frontiers of Foreign Language Education Research*, 3(3), 42 – 49.
- [23] Zhao, X. Y. (2024). A study on multimodal teaching models for English stylistic writing in the digital era. *Journal of Tongren University*, (02), 36 – 41.