rEFLections
Vol 31, No 2, May - August 2024

Students' Perceptions of Intersemiotically Cohesive Teaching Materials Designed for English for Agriculture

MAREEYADAR TEEDAAKSORNSAKUL
Faculty of Liberal Arts, Maejo University, Thailand
NEIL E. J. A. BOWEN*
Faculty of Liberal Arts, Thammasat University, Thailand
Corresponding author email: nbowen@tu.ac.th

Article information
Article history:
Received: 26 Oct 2023
Accepted: 9 Aug 2024
Available online: 13 Aug 2024

Keywords:
Multimodal material design
Design-based research
Intersemiotic cohesion
ESP
Classroom intervention

Abstract

Multimodal materials have become ubiquitous in language teaching classes. However, growing evidence suggests that misalignments between multimodal elements can mislead or even confuse students. To explore such concerns, this paper focuses on the design and pedagogical application of intersemiotically cohesive multimodal materials in an English for Specific Purposes (ESP) context. Specifically, utilizing a classroom-based intervention with two groups of Thai undergraduates (N = 41) enrolled on an English for agriculture course, we employed think-aloud protocols and video recordings to capture their interactions with multimodal materials we designed. The video recordings were then used for video-stimulated focus group interviews with six representatives from each group. This allowed us to explore how these students perceived and interacted with our designed materials and enabled us to investigate how a congruence between written text, visuals, hyperlinks to translations and audio, and consciousness-raising questions affected their learning experience. The findings revealed that well-integrated multimodal resources supported vocabulary acquisition and reading comprehension. However, students reported challenges with the abstraction of text and usability of hyperlinks, suggesting areas for further material refinement. Based on our results, we make recommendations for multimodal material design and underscore the critical need for cohesive educational resources to ensure learning effectiveness and accessibility.

INTRODUCTION

The term "multimodal literacy" emerged in the late 1990s following the work of the New London Group (1996). In language learning settings, multimodal literacy can help students understand and learn the target language by exploiting semiotic resources besides language alone (Guichon & McLornan, 2008; O'Halloran et al., 2016; Sueyoshi & Hardison, 2005). A multimodal approach to learning can also raise students' cultural awareness (Busà, 2010; 2015) and motivation (Shih, 2014). However, concerns have been raised about how possible disjunctions in meaning when individual modes are combined can lead to confusion or misunderstandings among learners (Bateman, 2014; Engebretsen, 2012). For instance, weak

semantic ties between visual and linguistic elements (see Bowen & Teedaaksornsakul, 2024) can lead to misalignments (or tensions) between the meanings conveyed by each modality, as well as the overall meaning(s) conveyed by the combined form. It is sometimes argued that this can lead to comprehension challenges for those consuming such content (Engebretsen, 2012; Fadel, 2008).

A key theoretical development in this regard has been the notion of *intersemiotic cohesion* (Liu & O'Halloran, 2009). Intersemiotic cohesion refers to how elements from different modes of communication interact with each other to create a cohesive whole. Thus, intersemiotic cohesion builds upon the linguistic view on cohesion, where elements such as nouns and their pronouns, close synonyms, and grammatical markers contribute to a text's cohesion (see Halliday & Hasan, 1976), and considers how links is achieved between elements in a multimodal text. Fundamentally, intersemiotic cohesion reflects how different modes (e.g., written text, images, and sounds) complement, enhance, or contradict each other when placed together to form a multimodal text. Using the concept of intersemiotic cohesion, educators and scholars have focused on developing multimodal design approaches to materials and pedagogy (Guichon & McLornan, 2008; Sueyoshi & Hardison, 2005), illustrating how the integration of different modes of meaning can impact learning experiences, rendering them more or less engaging for learners (Kalantzis & Cope, 2023; Kress & Selander, 2012; Lim et al., 2022a). The main argument here is that the different modes should be integrated in a coherent, meaningful, and unified way. Indeed, when the contributions of different modes are in unison, learning materials are said to be more engaging and easier to understand than when modalities represent conflicting or disconnected meanings (Bowen & Teedaaksornsakul, 2024; Kress & Selander, 2012). However, to date, no study has explored the application of such an approach in designing and testing multimodal material that is intersemiotically cohesive from the perspective of English for Specific Purposes (ESP) students, for whom such material can be instrumental in their learning goals (Bowen & Teedaaksornsakul, 2024).

Accordingly, this study reports on Thai undergraduate ESP students' perceptions toward multimodal material that was designed to be intersemiotically cohesive. More specifically, through think-aloud protocols and video-stimulated recall focus group interviews, we explore the student's obstacles and achievements when using visual-verbal interfaces (combining image and written text), hyperlinks to audio and definitions/translations, and consciousness-raising (C-R) questions designed for an English for agriculture course. Ultimately, by bringing together theory and practice, we hope our study can further clarify how to design and apply intersemiotic cohesion in multimodal materials for ESP contexts and multimodal literacy in general. The study addresses the following questions:

(1) How do Thai undergraduates studying on an ESP agriculture course perceive the practicality and effectiveness of intersemiotically cohesive multimodal resources?

(2) How can the challenges they encounter, if any, inform future multimodal material design in an ESP context?

LITERATURE REVIEW

The impact of technology and multimodality on modern educational materials

Multimodality significantly shapes the design and content of modern-day textbooks, reflecting a crucial shift in how educational materials are crafted to enhance learning outcomes. Lim et al. (2021), for example, state, "Literacy is no longer a question of reading and writing language-based texts; students also need to develop multimodal literacy in response to the communicative demands of the digital age". (p. 730) In other words, the increasing use of multiple semiotic modes necessitates corresponding multimodal pedagogies to more accurately reflect modern-day learning needs (Bezemer & Kress, 2016). Indeed, extensive research on multimodal teaching and learning asserts that effective multimodal design can engage students in lessons, leading to improvements in learning, memory retention (Mayer, 2014), understandings of a target language, and English skills (Yulini, 2022).

However, while the potential benefits of multimodal learning environments are widely acknowledged, there is a need for a more nuanced examination of learning outcomes across various educational contexts. Consequently, educational researchers have shown an increasing interest in multimodal materials design. Specifically, different learning materials, such as auditory, computer-based, and web-based blended learning, are being developed to facilitate the teaching of the English language (Tomlinson, 2023). Additionally, multiple studies indicate evident changes from predominantly text-based printed textbooks with pictures (e.g., Jewitt, 2006; O'Halloran, 2005) to digital formats with changes in layout and placement of semiotic resources (e.g., Bezemer & Kress, 2010; Kress & van Leeuwen, 2020; Motteram, 2011), and a corresponding shift from publications primarily created by established publishers to those created by independent developers (Motteram, 2011). These changes signify a broader trend toward more flexible and adaptable educational tools that are better suited to modern learning contexts and digital literacy demands.

In the context of materials for learning English as a foreign language (EFL), an increase in visuals and more elaborate layouts has also been found. For instance, Yassine (2014) observed a transition in the layout design of three localized Algerian EFL textbooks to align with Algerian educational standards for secondary students, moving from the conventional separation of text and image to different layout styles, including integration (text and picture coexist in the same space, with the text seamlessly integrated and superimposed on the landscape scene), overlap (a porous frame with parts of the picture breaking through, letters overlapping into the pictorial space), and rhythm (referring to two distinct elements sharing a common feature). Moreover, researchers have identified changes in the types of images used, for example, shifting from cartoons to photographs in more recent textbooks (Chen, 2010). These changes in visual structures have contributed to variations in text styles based on educational levels, where the presentation of material often progresses from narrative representations for lower levels to conceptual representations for higher levels, the latter of which draw heavily on reporting and expounding (Guo & Feng, 2015; Liu & Qu, 2014).

However, the shift from traditional to digital formats has not only changed the design of educational materials but has also raised questions about the pedagogical efficacy of these

designs (Bezemer & Kress, 2010; Motteram, 2011). For instance, the assumption that more dynamic and visually engaging materials automatically leads to better learning outcomes is overly simplistic, and it perhaps overlooks critical factors such as the cognitive load imposed by complex multimodal designs and their usability. Furthermore, the move towards materials created by independent developers, as noted by Motteram (2011), introduces variability in quality and accessibility. Thus, while a democratization of content creation can bring innovation and customization, it also requires rigorous evaluation to ensure educational standards are maintained.

The necessity of effective pedagogical strategies when using multimodal materials

The presence of multimodal material in English language teaching (ELT) has been a long-standing affair (Coccetta, 2018; Guo & Feng, 2015). However, as noted in the previous section, the affordances of new technologies that drive quick and easy content creation challenges us to (re)examine how modern-day materials incorporate diverse semiotic modes such as text, videos, and features. Fundamentally, with access to technologies such as generative AI and easy-to-use editing software (e.g., Canvas), any teacher with a computer can now design and publish multimodal materials relatively easily and quickly. Consequently, as argued by Bezemer and Kress (2010) over a decade ago, there is a real need, now more than ever, for a deeper understanding of how meanings can be collaboratively constructed through multiple modalities: The crux of the argument being the need for educational materials that not only present information, but do so coherently and in a cohesive manner, where pedagogical efficacy and applicability to a wide range of learning situations is also important (Danielsson & Selander, 2016; Hadjiconstantinou, 2021; Lim et al., 2022a).

This need for a clearer understanding of how modalities interact in educational materials aligns with increasing scholarly attention to intersemiotic relations among different semiotic modes, such as visuals and verbal (or language-based) texts (Elmiana, 2019; Guo & Feng, 2015; Pertama et al., 2018). Nevertheless, despite theoretical advances in comprehending these complex interactions, the implementation of such theory into practice has been slow. For example, studies by Derakhshan (2021), Elmiana (2019), and Mayer (2014) have all observed that visuals in ELT textbooks are frequently used for decorative purposes rather than meaningful, integrative tasks. However, as noted by Liu and Qu (2014), Sidabutar et al. (2021), and Yassine (2014), while visuals can enhance and supplement textual content, there is a real need in materials design to transition from mainly aesthetic appeals to functional affordances.

Such concerns also highlight the importance of intersemiotic cohesion, which can be a key contributor to developing coherent and integrated materials (Royce, 2002; Sanchez-Stockhammer, 2021; Schubert, 2021). For example, Attar (2014) and Khalid et al. (2017) investigated how different semiotic resources contributed to text coherence in educational materials and found that multimodal texts with lower cohesion hindered students' understanding and learning. Khalid et al.'s (2017) research found that consequence relations, in which one semiotic mode represents a cause while the other mode shows its effect, were too complex for elementary students to grasp. Liu and Qu (2014), on the other hand, propose that verbal texts establish cohesive relations with accompanying visuals to convey meanings

for various picture styles, such as narrative, conceptual, and abstract, catering to different English proficiency levels. Drawing on Liu and O'Halloran's (2009) framework of intersemiotic cohesive devices, these studies illustrate how strategies, such as intersemiotic parallelism, seek to create a deeper "intersemiotic texture" inside texts, pushing beyond mere connection and toward a more thorough integration of words and images (see also Bowen & Teedaaksornsakul, 2024). This integration, in turn, helps to move a text beyond simply presenting information, and engenders material that can provide support and facilitate the understanding of language content (see also Salbego et al., 2015).

Of note here, is a recent book chapter by Bowen and Teedaaksornsakul (2024), who show the benefits of bridging the gap between multimodal theory and design practice to create intersemiotic cohesion in visual-verbal interfaces. Drawing on Halliday's Systemic functional grammar (Halliday & Matthiessen, 2014) and Liu and O'Halloran's (2009) concept of intersemiotic cohesion, Bowen & Teedaaksornsakul (2024) illustrate how cohesive visual-verbal interfaces can be designed to represent commensurate experiential meanings: The visual component conveys experiential meanings through narrative configurations, which use graphics to represent events over time and illustrate the actions and relationships of individuals involved; conceptual configurations, on the other hand, classify and interpret phenomena using graphs, diagrams, or pictures to organize or establish relationships between concepts. Meanwhile, the verbal (or written text in this instance) can convey commensurate experiential meanings through choices in Transitivity that align with the meanings construed by narrative or conceptual image configurations (and vice versa). Transitivity choices are realized by process types (main verbs in traditional grammar), participants (Complements), and circumstances (Adjuncts). The intersemiotically cohesive materials used in the current study were designed on this basis, further details of which can be seen in Bowen and Teedaaksornsakul (2024).

Pedagogical strategies for multimodal materials

While many studies, including those by Coccetta (2018), Lim et al. (2021), Kress and Selander (2012), and Morell et al. (2020), highlight the importance of multimodal pedagogy in developing learners' multimodal competence, the mere presence of intersemiotically cohesive material does not guarantee effective learning. Essentially, as Lim (2018) contends, without the active use of multimodal pedagogy, students may still struggle to fully comprehend the meaning-making processes provided by semiotic resources, potentially leading to misunderstandings. Consequently, the urge for instructors to develop and possess multimodal literacy skills is not merely a recommendation, but a requirement. Indeed, current research suggests that teachers must be skilled at developing multimodal texts that are not only intersemiotically cohesive but also improve comprehension—elements that are critical for scaffolding learners' knowledge of multimodal interactions in language learning (Lim & Tan-Chia, 2023). However, as Kress and Selander (2012) highlight, although multimodal pedagogy can scaffold language acquisition by illustrating the interconnectedness of various modes, effective implementation of such strategies necessitates a deliberate and carefully planned alignment of semiotic resources with pedagogical goals.

One such strategy to align multimodal materials with pedagogic goals involves the use of questioning methods, as highlighted by Lim (2018), McDonald (2014), and Misa (2023). These

researchers, and others, argue that the use of consciousness-raising (C/R) questions can enhance learners' interactions with multimodal texts; nevertheless, the effectiveness of such tactics is heavily dependent on the learners' current competencies and the context in which they are used. For example, while Misa (2023) shows that engaging inquiries might improve the perception of representational visuals, there is a risk that students who lack appropriate background knowledge or contextual understanding will find these questions confounding rather than clarifying. Furthermore, McDonald's (2014) use of multimodal C/R questions, while unique, emphasizes the importance of educators receiving comprehensive training in both the theory and practical application of visual grammar, as advocated by Kress and van Leeuwen (2006). Such concerns underline a greater difficulty in multimodal pedagogy: the need for careful tailoring of teaching tactics to student capacities and learning settings. Nevertheless, for practical reasons, this study employs McDonald's (2014) multimodal C/R questions as a way for participants to critically engage with our intersemiotically cohesive materials, as these questions have been shown to be well-suited for exploring how semiotic resources interact and encourage active cognitive engagement.

Stakeholder perspectives and the role of design research in multimodal ELT pedagogy

Of particular interest to our research goals are those studies which have investigated the experiences and perspectives of stakeholders, teachers, and learners on the use of multimodal texts (e.g., Trisanti et al., 2022) and multimodal pedagogy (e.g., Haryyadi & Rohmah, 2023). These studies have shown that multimodal texts enhance motivation, comprehension, and engagement, promoting student engagement, teamwork, inclusivity in classrooms, and accommodating diverse learning styles. However, these studies are largely concerned with general EFL contexts and thus may not be directly transferable to specialized settings such as English for ESP contexts. In this specialized setting, learners are frequently exposed to content that is designed to meet occupational or professional needs, which could necessitate a different approach to multimodal pedagogy (Camiciottoli & Fortanet-Gomez, 2022; Laadem & Mallahi, 2019). Additionally, while these EFL studies provide valuable insights, their methodologies have limitations. For instance, studies like Giron-Garcia and Fortanet-Gomez (2023) and Haryyadi and Rohmah (2023) rely on self-reported data collected primarily through questionnaires, interviews, and micro-teaching without triangulating with actual classroom observations. Furthermore, Giron-Garcia and Fortanet-Gomez (2023) mostly relied on ESP teachers' self-reported data about how they chose multimodal videos and how they felt about using them. Hence, there were no direct observations of students' perspectives and interactions with the materials.

Going beyond simply evaluating multimodal texts and pedagogy from the perspectives of stakeholders, some studies illustrate how effective materials design involves iterations, collaborations, and evaluations with end users (Lim et al., 2015; Lim & Nguyen, 2022; O'Halloran et al., 2012). These studies, which aim to bridge the gap between multimodal design theory and practice, employ a design research approach, which provides "a rigorous framework for designing real-world solutions … while allowing for creativity and flexibility within the framework." (Archer, 2019, p. 320) For example, O'Halloran et al. (2012) designed multimodal analysis software that allows for the annotation, analysis, searching, and retrieving of semantic patterns

in complex semiotic modes, such as the interaction of intonation, music, camera angle, gesture, and gaze in a video. In a follow-up study, Lim et al. (2015) used this software to explore teachers' and students' experiences with a multimodal-based classroom intervention. They took a comprehensive approach to teaching visual texts, giving pupils meta-language and tools for evaluating multimodal texts. Feedback suggested that these resources increased engagement among learners and made learning more dynamic and entertaining. Some students, however, suggested that simpler interfaces and more training be provided to properly utilize the software's possibilities. In a similar study, but on a larger scale, Lim et al. (2022b) conducted design-based research that involved multimodal lessons and assessment packages. Through semi-structured interviews with teachers and students, they found that the designed lesson packages helped to improve students' performances and fostered teachers' professional growth.

Overall, while studies such as those listed above have provided valuable information on stakeholder experiences with and perceptions of multimodal approaches, so far, no study has explored the practicality and effectiveness of intersemiotically cohesive materials in an ESP classroom setting. Therefore, the current study, which is part of a broader design research project, aims to fill this gap by gathering ESP students' perspectives during and after their interactions with intersemiotically cohesive materials. The research also explores the use of multimodal C/R questions as a teaching approach to supplement such materials and thus enhance English language learning.

METHODOLOGY

Research design

This study reports on the assessment phases of a larger PhD project that uses educational design research as its underlying framework. Educational design research is "the systematic study of designing, developing and evaluating educational interventions" (Plomp, 2013, p. 9). It primarily involves preliminary research, a prototype phase, and an assessment phase. According to Plomp (2013), the developmental phase or prototype phase involves identifying necessary improvements, redesigning an intervention, and then trying out and revising these improvements based on formative evaluations such as expert judgments. The assessment phase, meanwhile, aims to determine "whether the solution or intervention meets the pre-determined specifications" (Plomp, 2013, p. 15).

The assessment phase we report on in this study involves a classroom intervention using intersemiotically cohesive materials and C/R questions. The materials were also localized to the specific context and topic by drawing on the findings of a needs analysis assessment. The design process and needs analysis are outlined elsewhere (see Teedaaksornsakul, 2024). Given that the course objectives focus on vocabulary, reading, and listening in agricultural contexts, and following the results of the needs analysis that revealed a limited desire for speaking and writing skills (Teedaaksornsakul, 2024), the predetermined specifications for the intervention were as follows:

• In terms of vocabulary enhancement, both still and moving images were designed to represent key vocabulary terms. Conceptual images were used to depict nouns and adjectives, while narrative or moving images that exhibited a similar transitivity structure were used to represent verbs.

• In terms of pronunciation and word meanings, hyperlinks to audio sounds of the vocabulary were created. Additionally, definitions of the vocabulary were provided, with hyperlinks to Google Translate for only challenging words within the definition sentences.

• In terms of listening skills, images were designed to represent the main ideas of the dialogues to scaffold students' comprehension of the dialogue's context before listening to the conversation. This was intended to help students fill in the missing words in the provided gaps within the dialogue scripts effectively.

Research context

The study took place at a large university in Northern Thailand, and it is part of a larger project aimed at designing and trying out a multimodal textbook tailored for Thai agricultural undergraduates enrolled on a 15-week ESP course. The course's learning objectives focus on recognizing subject-specific vocabulary and grammatical structures, applying these in specific contexts, and interpreting agricultural content primarily through reading and listening skills. At the time of this study, the course was administered online to 5 sections, with an average of 43 students per class; however, only two of these sections—Group 1 with 26 students and Group 2 15 students—contained agricultural students, who were the primary focus of this study.

Participants

The study involved a convenience sample of 41 Thai agricultural students from the academic year 1/2023. These participants, primarily second-year students along with some third- and fourth-year students retaking the course, were from two sections. All 41 students were exposed to the designed materials. Second-year students were specifically chosen because they are at a critical point in their academic journey where they have acquired foundational knowledge in their field but are still in the process of developing more advanced skills. This makes them ideal candidates for evaluating the effectiveness and practicality of the multimodal materials designed for the ESP course. Additionally, second-year students typically have a clearer understanding of their learning needs and challenges compared to first-year students, while not yet being as specialized as third or fourth-year students. This balance allows for a more representative assessment of the materials' impact on learners who are transitioning from basic to more complex subject matter within their curriculum. Including second-year students rather than those in other years also aligns with the course's structure and learning objectives, which are designed to build on the foundational knowledge acquired in the first year. By focusing on second-year students, the study ensures that the participants are engaged with the content at an appropriate level of difficulty, providing more relevant and actionable feedback on the multimodal materials.

From each group, six second-year students (n = 12) participated in follow-up video-stimulated recall focus group interviews. The first group consisted of three females and three males, and the second comprised two females and four males. The average age of these students was 19. Their English proficiency levels ranged from A1 to A2, as assessed through a standardized test. The selection of these 12 students was influenced by several factors. First, the specific objectives of investigating in-depth perceptions and thorough interactions with ESP-focused multimodal materials necessitated a small, focused sample. A smaller sample size is often advantageous in qualitative research as it enables a deeper exploration of individual experiences and provides rich, detailed data that can reveal underlying patterns and insights. Thus, our approach allowed for a detailed qualitative analysis using think-aloud procedures and video-stimulated recall interviews, which gather precise and nuanced information about student experiences and reactions.

Procedure

The first phase of data collection took place during one online class period (two hours). Six volunteer representatives from each group (12 in total) were asked to video record themselves in Thai (L1) while using the designed material and to think-aloud while doing so, a useful technique for studying user interactions and evaluating their effectiveness (Cotton & Gresty, 2006). Four main tasks were completed within the class period, with the researcher facilitating the students' access to the materials, helping as and when needed, and taking notes on the students' challenges and successes. This method emphasized feedback on the designed material for scaffolding English language learning rather than determining students' language skills and provided a rich data set of student interactions with the material.

During the think-aloud protocol, the students were asked to engage in four activities:

> (1) Pre-reading: Students read and answered three multimodal C/R questions, prompting them to (a) identify the salient visual elements to discern the conveyed meaning; (b) identify the events or concepts represented; and (c) infer the main idea of the image.
>
> (2) Previewing keywords: Students were challenged to learn the keywords of the reading passage by examining the relationship between the interwoven multimodal resources: visuals, written texts, hyperlinks to L1 translations, and hyperlinks to audio sounds illustrating how the keywords are pronounced.
>
> (3) Reading: Students were instructed to read a passage, examine the accompanying images, and fill in the blanks with the correct items and choose a correct answer between A or B that uses the underlined part correctly.
>
> (4) Listening: Students were asked to listen to a conversation and answer a question, analyse an image to find the main idea of the conversation, and then listen again to complete the provided gaps in the dialogue.

After completing the think-aloud process, the twelve students were asked to review their respective video recordings before participating in online stimulated-recall interviews; this was done to mitigate any potential problems of fatigue and boredom during the stimulated-recall focus group interview. The interviews were done on the same day as this approach can help participants recall their experiences more clearly (Paskins et al., 2017). Interviews were administered by the first researcher, who has taught the ESP course in questions and is an L1 Thai squeaker and experienced English language teacher, giving us an insider perspective on the data.

Data coding and analysis

This research used evaluation coding based on Saldana's (2014) coding methods (utilizing – and + symbols to indicate negative and positive evaluations, together with In Vivo or descriptive codes during the coding process). The evaluation coding was used to investigate the participants' evaluations of the designed multimodal resources (Table 1). A member-checking process was conducted to maximize the trustworthiness of the findings, wherein participants were invited to check the interpretation of the findings.

Table 1

Examples of coding schemes (based on Saldana, 2014)

Theme	Sub-theme	Codes	Examples of quotes
Usefulness and effectiveness of materials in scaffolding vocabulary learning	Providing concrete representation and context cues of the word	+ INTEGRATED SEMIOTIC RESOURCES: "EASY TO USE"	"It is also easy to use and familiar to us, so we can learn easily".
		+ INTEGRATED SEMIOTIC RESOURCES: "EASY TO UNDERSTAND"	"... the picture of the growing plant was easy to understand and could analyse what it meant".

The data from the 12 participants were collectively analyzed to gain a comprehensive understanding of student experiences and perceptions across both sections. This method enhanced the robustness of our findings by consolidating data from both focus groups.

RESULTS AND DISCUSSION

This section seeks to answer our first research question, and thus explores how the sampled Thai undergraduates perceived the practicality and effectiveness of the multimodal resources.

Usefulness and effectiveness of C/R questions in scaffolding reading skills

The focus group interview results and our observations of the think-aloud protocols revealed two main benefits of using C/R questions: (1) as a priming activity before reading, and (2) as a scaffolding device for identifying main ideas.

Priming function of multimodal material

Ten of the 12 students remarked on the usefulness of the multimodal C/R questions in priming them for the upcoming learning activity. For example, Student 5 noted that the multimodal C/R questions "helped prepare me [sic] to understand the reading passage". Student 3 explained that the question prompted her to "see what would be in the lesson … they are basic questions that help us to know what we will learn in our lesson". Student 2 added, "[the questions] helped us observe the picture so that we knew about the topic we would study … helped us understand the content we will read next". Student 9 also added that they perceived the multimodal C/R questions as a "necessary" tool that helped students "notice the key idea of the images before reading". Student 6 specified, "They helped us observe the picture and look at the words to let us know what words we need to use in the lesson". Student 10 concluded that the multimodal C/R questions helped her "observe the pictures" and "draw attention and motivate me to learn".

These reflections are consistent with the findings of many studies that assert the advantages and impacts of visual images on the learning process, as they help prepare learners for activities or exercises within a textbook (e.g., Liu & Qu, 2014; Royce, 2007; Weninger, 2021). These echo, for example, Royce's (2007) assertion that the efficiency of previewing images before reading can help build confidence and make learners comfortable with the reading process as they "get some idea of what to expect ... The effect is that expectancies are being set up in the students' minds, and the process of reading the text will then either give them a confirmation of their interpretation of the information (or of the story)" (Royce, 2007, p. 380), resulting in effectively scaffolding reading skills.

Scaffolding reading skills

Regarding scaffolded learning, most participants (n = 10) stated that the multimodal C/R questions were helpful and practical in identifying the outstanding visual elements within the pictures, which helped them to discern the main ideas of the reading section. For example, the following excerpts show how the multimodal C/R questions are helpful and practical:

> They [= the multimodal C/R questions] helped us observe things by looking at the given words, and the pictures helped us find the main point. (Student 4)
>
> I think it's effective. The question asked us to look at what the story is about. I think it helped us see the main point of the story… It's useful. (Student 6)
>
> The visual elements help us notice the picture and see what the people in the picture are doing… Observing a picture's visual elements and details can help us understand what it is about. (Student 11)
>
> They help us notice and understand what the story will be about. (Student 12)

These remarks show how C/R questions can draw students' attention to salient visual elements that are intersemiotically parallel to the reading passage's central idea, helping them to understand the central message. This finding highlights the importance of Liu and O'Halloran's (2009) concept of Intersemiotic Parallel Structures by revealing how multiple modes can effectively convey the same meaning through shared structures. Eight students commented on how the designed images, which aligned with the content, were useful and enhanced comprehension. For example, Student 11 remarked, "The images of Thai plants for each season shown in the calendar, representing each month with symbols of seasons clearly separated by the table, are useful." Then, Student 7 elaborated, "This helps us know which season is suitable for planting or harvesting them." Additionally, two other students admitted the benefits of multimodal learning aids. Student 3 said, "My English was very poor, but the moving image of the woman picking vegetables alongside the vocabulary helped me understand the meaning of the word 'harvest'." Meanwhile, Student 5 explained, "It's easier to understand when there are both texts and images." Overall, the findings underline the significance of co-contextualization and the convergence of textual elements in enhancing understandings.

Usefulness and effectiveness of intersemiotically cohesive resources in scaffolding vocabulary learning

The results revealed three main benefits of using the intersemiotically cohesive materials: (1) as a tool to clarify word meanings, (2) as a quick and easy way to retrieve direct translations, and (3) as a pacing mechanism for learning.

Providing concrete representation and context cues of the word

Most students expressed that using integrated multimodal resources was useful and practical for vocabulary learning, as the combined materials were able to fully represent the meaning of the words and give contextual cues. For example, Student 7 remarked, "The semiotic resources are helpful, especially for vocabulary learning". Student 11 added that he succeeded in comprehending "most of the vocabulary … as they [= the materials] conveyed the meaning, images, and sound, which made me understand and learn vocabulary better". Student 1 further explained how the integration between the semiotic resources provided concrete representations of the words, enabling learners to comprehend the conveyed meaning better and enhance the learning experience: "the moving image of the word 'harvest' helped me differentiate its meaning as a verb and a noun. The image of the plants in the containers helped me understand that someone collected the harvests."

The perception that using intersemiotically cohesive resources can enhance vocabulary learning confirms its importance when designing language learning material. Fundamentally, intersemiotic cohesion can scaffold English language learning, as such resources enhance learners' comprehension of the conveyed meaning and content, respectively; this is consistent with the findings of Salbego et al. (2015) and Khalid et al. (2017), who showed how aligning visual and verbal texts in English textbooks significantly aids beginners' comprehension as this process works by facilitating logical connections between the visual and verbal elements. It also echoes the views of Bogaards and Laufer (2004), Milton (2009), and Nation (2001),

who argue that vocabulary is best learnt through exposure to multiple modes. Indeed, using visuals as a priming activity is a common tactic among experienced language teachers in contexts like ours (Thomas et al., 2023).

Using hyperlinks to provide translations

Of the three multimodal resources—visual elements, hyperlinks to audio sounds, and hyperlinks to Thai translations—participants' feedback indicated that each were perceived as equally useful and necessary; nevertheless, there was a slight preference towards hyperlinks to Thai translations, as expressed by Student 8: "Apart from the images, I like the hyperlinks to Google Translate. It would be more helpful if you created the hyperlinks to the translation of the vocabulary directly." Other students echoed this desire; for instance, Student 12, who considered the hyperlinks to translation as necessary, suggested, "They did not represent many words, which would be great to have more of them". This might be attributed to the fact that the hyperlinks to the dictionary and Google Translate were only included for keywords within the texts, providing definitions of key vocabulary and a partial representation of the multimodal resources conveying the meaning. However, novice learners might require additional translations of the words provided in the definition text to assist them in accurately deducing their meanings.

The request for more hyperlinks to Google Translate even though the reading passage's topic was quite general, seems to reflect the participants' status as novice English language learners. Essentially, their remarks indicate that they are still dependent on first language translations as they struggle to guess meanings despite being given additional contextual information regarding the word's usage (i.e., visual representations). This is in line with the findings of Almusharraf and Bailey (2023) regarding the preference of EFL learners for using Google Translate as a tool for improving their English language skills, and Yu and Liu (2022), who found that the use of L1 translation together with other semiotic resources offered more benefits than drawbacks. However, the inclusion of L1 translation as a learning strategy remains a contentious issue (Aithal, 2023). According to Yu and Liu (2022), for instance, while the use of L1 translation is necessary to help learners acquire new vocabulary, the choice of whether to make it accessible to students either before or after other semiotic resources is contingent on individual learning styles, as some learners benefit more from L1 translation, while others may prefer images, depending on the context.

Moreover, our findings suggest that comprehending conveyed meanings may require more than just a visual image, especially for students who are not familiar with multimodal literacy practices and require adequate scaffolding. Indeed, in their research, Duke et al. (2013) argue that even when visuals are well-designed and of high-quality, novice learners who lack the skills to interpret them effectively will find it difficult to understand the information presented within the image. This may be due to an inability to distinguish between critical and non-critical information within the visual elements. (Duke et al., 2013). Ultimately, such issues emphasize the importance of establishing multimodal literacy practices to help scaffold the learning process, where coherent semiotic resources, including hyperlinks to the translation, can constitute valuable tools that should be both designed and employed according to individual leaner needs.

Taking an active role in one's learning

Some students remarked that the multimodal resources empowered them to take an active role in managing their learning, as the materials enabled them to control the pace of learning and improve their understanding. For example, Student 2 noted, "The images let me understand the concept of the words and the content, while the hyperlinks to audio sounds allowed me to listen to the conversation as many times as I needed". Student 11 had also the same impression, stating that "the semiotic resources helped me understand the content and remember the vocabulary better than listening to the translation from the teacher".

Unlike the others, Student 11 compared the benefit of using integrated semiotic resources for teaching and learning vocabulary with the traditional way of teaching and learning. Student 3 added that the multimodal resources "help me understand and remember the vocabulary better than when the teacher reads, translates, and plays an audio sound for us. I used to take notes, which was tedious and not very effective". These reflections suggest that multimodal text facilitates learning in different ways, engaging students more actively in the learning process and resulting in better retention. This is consistent with the findings of Chen and Jamiat (2023), who found that interactive multimodal learning of poetry improved EFL learners' comprehension by boosting their intrinsic motivation when they were allowed to learn independently at their own pace.

Overall, the participants' feedback on the multimodal resources was positive. Moreover, the results support the view that intersemiotically cohesive resources, particularly those that exhibit clear semantic links between visual elements and written text, can assist novice EFL learners in understanding the co-constructed meaning that they convey (Alyousef, 2021; Attar, 2014, Khalid et al., 2017).

Perceived challenges of the designed material

This section addresses our second research question by exploring the challenges of using intersemiotically cohesive material. Our goal here is to identify areas that need further improvement.

Understanding and visualizing abstractions

Although many students recognized the value and effectiveness of the multimodal resources in understanding the conveyed meaning of the content, particularly in the vocabulary and exercise sections, some participants encountered difficulties in grasping abstract words or concepts. Student 2 pointed out that" images and audio sounds helped me understand the vocabulary and the reading passage. However, I did not understand the arrow in the yearly calendar image. What does it tell us?"

Student 5 also pointed out, "some details in the reading passage were not represented. There should be more hyperlinks and images to represent more words, such as "while", which is abstract, to help learners better understand the meanings." These difficulties, as reported by

Students 2 and 5, may originate from either a deficiency in multimodal literacy or insufficient English language proficiency. Alternatively, it may be caused by a lack of clarity in the semiotic resources used to represent the idea.

Representing abstractions and connections precisely is inherently challenging, leading to ambiguity as these meanings can sometimes be interpreted in various ways (Bowen & Thomas, 2020). Furthermore, abstraction also points out the challenge of the resemiotization process. This process consists in transforming one semiotic mode to another to reduce the abstractness of complex, specialized content (Iedema, 2003). It is, however, important to remember that resemiotization does not imply a direct translation from one semiotic mode to another. When the mode of communication is altered, maintaining the original meaning becomes impossible, as each semiotic mode represents meaning differently (Iedema, 2003). Therefore, resemiotizing alone does not suffice and other semiotic resources, including simplifying words or content, should also be considered when designing multimodal text.

Overall, it is essential to not only employ multimodal design strategies to reduce abstractness, but also to simplify or reformulate the content, such as supporting processes such as nominalization through multimodal resources (Fernandez-Fontecha et al., 2019). Specifically, resources, such as visual aids, can help visualize the connections between nominalized terms and their characteristics. Furthermore, clear visuals and analogies that build on learners' previous knowledge can make abstract concepts easier to understand. Moreover, complex ideas should be broken down into smaller parts, each accompanied by visual aids, a simple explanation, and multimodal C/R questions to guide interpretation. Finally, understanding how to create intersemiotically cohesive materials is necessary for multimodal text designers, particularly for teachers who must tailor their materials to the specific needs and contexts of their learners (Teedaaksornsakul, 2024).

Relying on internet connections for hyperlinks

Some participants reported problems regarding the usability of the hyperlinks and issues with their accessibility. Some students, for example, had to deal with unstable internet connections, triggering difficulties in accessing the hyperlinks. Student 9 commented, "Sometimes the links were difficult to access. I had to log in to connect to the internet three times. It took a while to access the translation". Some other students shared the same concern, with Student 10 noting, "I experienced the same problem. I had to log in to my Google account to access the translation, which was inconvenient". Other participants pointed out that hyperlinks to the audio sounds could be active at the same time: Student 2 noted, "On the dialogue slide, all the sounds play at the same time". Student 8 shared the same experience and added, "some of the sounds overlap ... and while I was clicking on it, it was not accessible".

The challenges mentioned above highlight the nature of relying on an internet connection for accessing educational materials. While the internet, as a tool, can serve as "an enabler" of learning platforms, broadening and enhancing access to education (Mhlongo et al., 2023, p. 9), it can also act as a constraint when the connection is unreliable or unstable, leading to delays or interruptions in learning activities (Trisanti et al., 2022).

Understanding the need for prior knowledge in semiotic resource utilization

Although semiotic resources can effectively help visualize words or clauses, prior knowledge is required to fully understand the content. Indeed, a few students commented on the difficulty of completing one exercise despite the accompanying visuals representing the conveyed meanings of the statement. Student 4 remarked,

> "The images representing the vocabulary are clear … I knew that 'Tubers are used to make clothes' was incorrect, but I wasn't sure if 'Industrial crops are not eaten' was correct. The exercise seemed to test our knowledge rather than our understanding."

This impression was echoed by Student 6, who encountered a similar issue with a different exercise, "The second reading passage was difficult for me, although I could generally understand that there are two types of crop calendars from the images representing it". These reflections suggest that novice learners require extra semiotic resources to help with their comprehension of a text and background knowledge of the content being represented, as well as more prompts to scaffold their cognitive abilities. Also, these observations highlight the importance of additional semiotic resources and previous knowledge for novice learners' comprehension. Mayer (2014) notes that combining verbal and visual information can considerably improve learning, especially when learners can actively link this new information with their previous cognitive frameworks, which requires the use of prior knowledge. This integration process is critical because it allows students to draw meaningful connections between new material and what they already know, which improves retention and comprehension. Brod (2021) provides additional support for this idea. He argues that prior information has a major impact on learning outcomes, depending on how well it is activated, its relevance, and its congruence with the new context. This shows that effective learning requires active engagement of pre-existing knowledge structures rather than passive processing of information. Furthermore, Blagojević's (2013) research emphasizes the value of prior knowledge in the context of English for Specific Purposes (ESP), demonstrating that comprehension improves when students are familiar with the subject. This is especially important in ESP contexts, where content frequently includes specialized jargon and concepts that learners may not be familiar with.

To summarize, addressing the issues with intersemiotically coherent materials emphasizes the need for clearer representations of abstract concepts, alternatives to relying on stable internet connections, and effective ways for engaging learners' prior knowledge. Addressing these issues is critical to increasing accessibility and boosting learning results, ensuring that all students benefit equitably from educational resources.

Rethinking L1 translation: challenges or opportunities

Many students indicated the need for Google Translate or direct L1 translations, stating, "It is useful if you don't know what the word or sentence means" (Student 5). However, many students also reported difficulties accessing the links to Google Translate or complained that there were not enough links. Student 4 remarked, "I could not completely understand the

reading passage, so I need more hyperlinks to Google Translate." Student 2, meanwhile, expressed frustration, stating, "There were no hyperlinks provided for some words I needed translations for." Student 7 further elaborated, "There should be more hyperlinks for the details in the reading passage." This is consistent with other studies in similar contexts to ours, where students perhaps rely too heavily on L1 translations rather than strategies to guess meanings (e.g., Bowen et al., 2023). However, findings from numerous studies, such as those by Al- Musawi (2014), Calis and Dikilitas (2012), Pan and Pan (2010), and Yu and Liu (2022), all point to the perceived benefits of L1 translation in lowering anxiety, increasing confidence, and improving understanding among novice EFL learners.

However, an over reliance on L1 translation raises serious questions about the ultimate purpose of English language development. For example, while Pan and Pan (2010) observe that L1 translation can make learners feel more secure, and Al-Musawi (2014) notes its role in relieving anxiety caused by low English proficiency, reliance on L1 translation may inhibit immersion in the target language, potentially hindering the development of direct cognitive links with the English language. Nevertheless, Yu and Liu (2022) argue that combining L1 translation with multimodal learning materials, such as visual aids, can help scaffold language learning and improve vocabulary acquisition. However, this strategy could hinder students' ability to think and operate directly in English, as the continual availability of L1 support may lessen the need for deeper engagement with the English language. While this approach can be beneficial; it is crucial to maintain a balance. According to many scholars such as Littlewood and Yu (2011), Swain and Lapkin (2000), and Turnbull and Arnett (2002), they all emphasize the importance of using L1 strategically as a supplement to enhance, not hinder the learning of English.

Considering such issues, it is vital to evaluate how L1 translation is used alongside multimodal resources. Fundamentally, while hyperlinks to translations can be a useful tool for beginners, they should be gradually reduced as learners progress to encourage more direct connection with multimodal material used for English language. In other words, we believe that links to L1 translations should be reduced as learners progress, where the responsibility for understanding word meaning be shifted to intersemiotically cohesive materials. These materials should include strong visual scaffolding by matching visuals with concrete vocabulary that represents actions and characteristics. Moreover, establishing strong semantic ties between vocabulary and images is essential, as alignments between elements can enhance comprehension and reduce cognitive load for students (Bowen & Teedaaksornsakul, 2024). For instance, at the clause level, visual representations should display intersemiotic parallelism by linking text and image based on a shared transitivity structure (Liu & O'Halloran, 2009): An image illustrating a narrative should display similar participants (nominal or adjectival groups), processes (verbal groups), and circumstances (prepositional phrases, adverbial groups, etc.).

Materials such as those outlined above can, in turn, be supported with multimodal C/R questions to facilitate decoding. For reading passages, visuals that represent the main idea and major details should be designed to be intersemiotically cohesive and accompanied by C/R questions. This strategy could enhance understanding and reduce the need for L1 translations. If certain words or phrases cannot be effectively represented visually due to their complex transitivity structures, hyperlinks to translations may be provided. However, this

should be done selectively; hyperlinks should primarily target abstract concepts that are challenging to depict visually, such as metaphors or idioms. This careful selection ensures that the use of L1 translation is limited to instances where it is most needed, maintaining the focus on direct interaction with the English language. This balanced strategy would ensure that L1 translation promotes English language learning while not impeding the development of language competency and fluency.

CONCLUSIONS, LIMITATIONS, AND RECOMMENDATIONS

This study investigated ESP students' perceptions and feedback on using intersemiotically cohesive multimodal resources. These resources included written text, images, and hyperlinks to language support; together, these materials were design for an English for agriculture course at a Thai university, where we tested the effectiveness and usability of the designed materials through think-aloud protocols and video-stimulated focus group interviews. As such, we have made several contributions to the field of multimodal pedagogy and multimodal materials design, including empirically validating the effectiveness of multimodal resources in EFL learning, refining methodological approaches for studying their impact, supporting theoretical frameworks on multimodal literacy and scaffolding, and offering practical pedagogical recommendations for language educators.

Empirically, our findings contribute to the wider discourse on the use of intersemiotically cohesive materials in language learning. Specifically, our findings revealed that the participants perceived the materials as useful and effective in enhancing English language learning. This is perhaps not that surprising, However, our findings also show that when C/R questions are integrated with multimodal resources, they are effective in priming students and scaffolding reading skills. Specifically, students reported that the C/R questions helped them understand and anticipate content, thereby making the reading process smoother. Moreover, students found the integration of multimodal resources, such as visual elements and hyperlinks to translations, aided their vocabulary learning, as the materials provided concrete representations and contextual cues for new words. Furthermore, our findings indicate that well-designed multimodal resources can empower students to take an active role in their learning, allowing them to control the pace and better retain vocabulary and content. However, our findings also revealed how students experienced difficulties in understanding abstract concepts and using hyperlinks.

As for pedagogical recommendations, regarding the resemiotization of abstract concepts into different modalities, we recommend providing additional details or contextual information in callout boxes or drawing more on the affordances of multimodal C/R questions to guide learners in understanding abstractions. For issues surrounding the use of hyperlinks, we recommend that teachers do not rely too heavily on online translations when designing their primary materials, as these can lessen the efficacy of self-regulatory learning through discovery. Instead, we recommend a balanced approach to using L1 translations in multimodal material, advocating for its gradual reduction as learners progress to encourage direct engagement with the target language. Moreover, all provided materials, including hyperlinks, should be accessible offline to mitigate issues caused by poor internet connections, which are common in contexts like ours.

Theoretically, our findings support Liu and O'Halloran's (2009) concept of Intersemiotic Parallel Structures, demonstrating how multiple modes (text, visuals) can effectively convey the same meaning and enhance comprehension. The findings also reinforce the need for scaffolding in language learning by showing how multimodal resources can provide necessary support to help students, with which they can gradually achieve better understanding and autonomy in learning. Finally, our findings support the importance of fostering multimodal literacy, especially when integrating various semiotic resources to support language learning.

Our study has several limitations, First, we sampled only 12 Thai undergraduate students, which may not be representative of a broader population. Moreover, we focused specifically on an English for agriculture course, limiting the generalizability of the findings to other ESP contexts or general EFL learning environments. Thus, future studies would do well to extend data collection to include larger and more diverse samples (including teachers) as well as more varied learning contexts. Second, as with any study that relies on self-reported data, we acknowledge that students may have given biased or socially desirable responses that might not have accurately represented their true experiences and perceptions. To combat this, future studies may want to incorporate additional data sources into the design, such as pre- and post-measures of material learned, observational data, or even a control group that used non-cohesive multimodal material.

ACKNOWLEDGEMENT

This research was supported by Thammasat University's PhD scholarship for research of the academic year of 2021 and approved by The Human Research Ethics Committee of Thammasat University (Social Science) No.20/2022.

THE AUTHORS

Mareeyadar Teedaaksornsakul is a PhD candidate in English Language Studies at the Faculty of Liberal Arts, Thammasat University, and an English language instructor at Maejo University, Thailand. She has an MA in English from Chiang Mai University. Her research interests include semiotics, literary and textual studies, and English language teaching.
mareeyadar.maew@gmail.com

Neil E. J. A. Bowen is an associate professor at Thammasat University, Thailand. He is primarily interested in writing research, but has published on teacher identity, language learning strategies, and discourse analysis in TESOL Quarterly, the *Journal of Second Language Writing, Teaching and Teacher Education, System, Journal of Writing Research*, and other prominent journals.
nbowen@tu.ac.th

REFERENCES

Aithal, V. (2023, February 19). *When is it OK to use students' L1 in the English classroom?* Cambridge. https://rb.gy/tvhez

Al-Musawi, N. M. (2014). Strategic use of translation in learning English as a foreign language (EFL) among Bahrain University students. *Comprehensive Psychology, 3*, Article 4. https://doi.org/10.2466/10.03.IT.3.4

Almusharraf, A., & Bailey, D. (2023). Machine translation in language acquisition: A study on EFL students' perceptions and practices in Saudi Arabia and South Korea. *Journal of Computer Assisted Learning, 39*(6), 1988–2003. https://doi.org/10.1111/jcal.12857

Alyousef, H. S. (2021). Text cohesion in English scientific texts written by Saudi undergraduate dentistry students: A multimodal discourse analysis of textual and logical relations in oral biology texts. *SAGE Open, 11*(3), 1–13. https://doi.org/10.117/21582440211032194

Archer, E. (2019). Design research: Developing effective feedback interventions for school-based monitoring. In S. Laher, A. Fynn & S. Kramer (Eds.), *Transforming research methods in the social sciences: Case studies from South Africa* (pp. 317–336). Wits University Press. https://doi.org/10.18772/22019032750.25

Attar, M. M. (2014). *Inter-semiotic cohesion analysis of multimodal elements in Iranian English textbooks* [Doctoral dissertation, University of Malaysia], Universiti Malaya Research Repository. https://studentsrepo.um.edu.my/5414/

Bowen, N. E. J. A., & Teedaaksornsakul, M. (2024). Designing localized and intersemiotically cohesive material: Teaching agriculture through visual-verbal interfaces. In V. Lim & J. Pun (Eds.), *Designing learning with multimodality for English-Medium Instruction in Asia*. Bloomsbury.

Bowen, N. E. J. A., & Thomas, N. (2020). Manipulating texture and cohesion in academic writing: A keystroke logging study. *Journal of Second Language Writing, 50*, Article 100773. https://doi.org/10.1016/j.jslw.2020.100773

Bowen, N. E. J. A., Insuwan, C., Satienchayakorn, N., & Teedaaksornsakul, M. (2023). The challenge of teaching English writing in Thailand: A tri-ethnography of Thai university lecturers. *LEARN Journal: Language Education and Acquisition Research Network, 16*(2), 482–498. https://so04.tcithaijo.org/index.php/LEARN/article/view/266967

Bateman, J. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge.

Bezemer, J., & Kress, G. (2010). Changing text: A social semiotic analysis of textbooks. *Designs for Learning, 3*(1–2), 10–29. https://doi.org/10.16993/dfl.26

Bezemer, J., & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. Routledge.

Blagojević, S. (2013). Original texts as authentic ESP teaching material – The case of philosophy. *ESP Today, 1*(1), 113–126. https://www.esptodayjournal1.org/pdf/ vol11/7.%20SAVKA%20BLAGOJEVIC-%20full%20text.pdf

Bogaards, P., & Laufer, B. (Eds.). (2004). *Vocabulary in a second language: Selection, acquisition and testing*. Benjamins.

Brod, G. (2021). Toward an understanding of when prior knowledge helps or hinders learning. *npj Science of Learning, 6*, Article 24. https://doi.org/10.1038/s41539-021-00103-w

Busà, M. G. (2010). Sounding natural: Improving oral presentation skills. *Language Value, 2*(1), 51–67. https://repositori.uji.es/xmlui/bitstream/handle/10234/28328/Maria_Grazia_Busà.pdf

Busà, M. G. (2015). Teaching learners to communicate effectively in the L2: Integrating body language in the students' syllabus. *Lingue e Linguaggi, 15*, 83–98. https://doi.org/10.1285/i22390359v15p83

Camiciottoli, B. C., & Fortanet-Gómez, I. (2022). Towards developing multimodal literacies in the ESP classroom: Methodological insights and practical applications: Introduction to the SPECIAL ISSUE Multimodal approaches in ESP: Innovative research and practice. *Multimodal Communication, 11*(1), 1–4. https://doi.org/10.1515/mc-2021-0021

Calis, E., & Dikilitas, K. (2012). The use of translation in EFL classes as L2 learning practice. *Procedia - Social and Behavioral Sciences, 46*, 5079–5084. https://doi.org/10.1016/j.sbspro.2012.06.389

Chen, C., & Jamiat, N. (2023). A quantitative study on the effects of an interactive multimodal application to promote students' learning motivation and comprehension in studying Tang poetry. *Frontiers in Psychology, 14*, Article 1189864. https://doi.org/10.3389/fpsyg.2023.1189864

Chen, Y. (2010). Exploring dialogic engagement with readers in multimodal EFL textbooks in China. *Visual Communication, 9*(4), 485–506. https://doi.org/10.1177/1470357210382186

Coccetta, F. (2018). Developing university students' multimodal communicative competence: Field research into multimodal text studies in English. *System, 77*, 19–27. https://doi.org/10.1016/j.system.2018.01.004

Cotton, D., & Gresty, K. (2006). Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology, 37*(1), 45–54. https://doi.org/10.1111/j.1467-8535.2005.00521.x

Danielsson, K., & Selander, S. (2016). Reading multimodal texts for learning – A model for cultivating multimodal literacy. *Designs for Learning, 8*(1), 25–36. http://dx.doi.org/10 .16993/dfl.72

Derakhshan, A. (2021). Should textbook images be merely decorative? Cultural representation in the Iranian EFL national textbook from the semiotic approach perspective. *Language Teaching Research, 28*(1), 79–113. https://doi.org/10.1177/1362168821992264

Duke, N. K., Martin, N. M., Norman, R. R., Knight, J. A., Roberts, K. L., Morsink, P. M., & Calkins, S. L. (2013). Beyond concepts of print: Development of concepts of graphics in text, preK to grade 3. *Research in the Teaching of English, 48*(2), 175–203. http://www.jstor.org/stable/24398654

Elmiana, D. S. (2019). Pedagogical representation of visual images in EFL textbooks: A multimodal perspective. *Pedagogy, Culture & Society, 27*(4), 613–628. https://doi.org/10.1080/14681366.2019.1569550

Engebretsen, M. (2012). Balancing cohesion and tension in multimodal rhetoric. An interdisciplinary approach to the study of semiotic complexity. *Learning, Media and Technology, 37*(2), 145–162. https://doi.org/10.1080/17439884.2012.655745

Fadel, C. (2008). *Multimodal learning through media: What the research says*. Cisco Systems.

Fernandez-Fontecha, A., O'Halloran, K., Tan, S., & Wignell, P. (2019). A multimodal approach to visual thinking: The scientific sketchnote. *Visual Communication, 18*(1), 5–29. https://doi.org/10.1177/1470357218759808

Giron-Garcia, C., & Fortanet-Gomez, I. (2023). Science dissemination videos as multimodal supporting resources for ESP teaching in higher education. *English for Specific Purposes, 70*, 164–176. https://doi.org/10.1016/j.esp.2022.12.005

Guichon, N., & McLornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36*(1), 85–93. https://doi.org/10.1016/j.system.2007.11.005

Guo, N. S., & Feng, D. (2015). Infusing multiliteracies into English language curriculum: The visual construction of knowledge in English textbooks from an ontogenetic perspective. *Linguistics and Education, 3*, 115–120. https://doi.org/10.1016/j.linged.2015.07.001

Hadjiconstantinou, S. (2021). Multimodal texts in support of linguistic and critical thinking in English for specific purposes. In S. Paapdima-Sophocleous, E. K. Constantinou & C. N. Giannikas (Eds.), *Tertiary education language learning: A collection of research* (pp. 29–45). Research-publishing.net.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

Halliday, M. A. K., & Matthiessen, C. N. I. M. (2014). *An introduction to functional grammar* (4th ed.). Routledge.

Haryyadi, J. C. A., & Rohmah, Z. (2023). Pre-service teachers' perceptions towards the implementation of multimodal texts in microteaching classes. *Jurnal Bahasa, Sastra, Seni, dan Pengajarannya, 51*(2), 255–269. http://journal2.um.ac.id/index.php/jbs

Iedema, R. (2003). Multimodality, resemiotization: Extending the analysis of discourse as multi-semiotic practice. *Visual Communication, 2*(1), 29–57. https://doi.org/10.1177/1470357203002001751

Jewitt, C. (2006). *Technology, literacy and learning: A multimodal approach*. Routledge.

Kalantzis, M., & Cope, B. (2023). Multiliteracies: Life of an idea. *The International Journal of Literacies, 30*(2), 17–89. https://doi.org/10.18848/2327-0136/CGP/v30i02/17-89

Khalid, S., Malik, M. A., & Butt, M. (2017). An Intersemiotic analysis of verbal and visual cohesion in primary-level textbooks. *Pakistan Languages and Humanity Review, 1*(1), 38–47. http://doi.org/10.47205/plhr.2017(1-I)1.4

Kress, G., & Selander, S. (2012). Multimodal design, learning and cultures of recognition. *The Internet and Higher Education, 15*(4), 265–268. https://doi.org/10.1016/j.iheduc.2011.12.003

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

Kress, G., & van Leeuwen, T. (2020) *Reading images: The grammar of visual design* (3rd ed.). Routledge.

Laadem, M., & Mallahi, H. (2019). Multimodal pedagogies in teaching English for Specific Purposes in higher education: Perceptions, challenges and strategies. *International Journal on Studies in Education, 1*(1), 33–38. https://doi.org/10.46328/ijonse

Lim, F. V. (2018). Developing a systemic functional approach to teach multimodal literacy. *Functional Linguistics, 5*, Article 13. https://doi.org/10.1186/s40554-018-0066-8

Lim, F. V., & Nguyen, T. T. H. (2022). Design-based research approach for teacher learning: A case study from Singapore, *ELT Journal, 76*(4), 452–464. https://doi.org/10.1093/elt/ccab039

Lim, F. V., & Tan-Chia, L. (2023). *Designing learning for multimodal literacy*. Routledge.

Lim, F. V., Cope, B., & Kalantzis, M. (2022a). A Metalanguage for learning: Rebalancing the cognitive with the socio-material. *Frontiers in Communication, 7*, Article 830613. https://doi.org/10.3389/fcomm.2022.830613

Lim, F. V., O'Halloran, K. L., Tan, S., & E, M. K. L. (2015). Teaching visual texts with the multimodal analysis software. *Educational Technology Research and Development, 63*, 915–935. https://doi.org/10.1007/s11423-015-9395-4

Lim, F. V., Toh, W., & Nguyen, T. T. H. (2022b). Multimodality in the English language classroom: A systematic review of literature. *Linguistics and Education, 69*, Article 101048. https://doi.org/10.1016/j.linged.2022.101048

Lim, F. V., Towndrow, P. A., & Min Tan, J. (2021). Unpacking the teachers' multimodal pedagogies in the primary English language classroom in Singapore. *RELC Journal, 54*(3), 729–743. https://doi.org/10.1177/00336882211011783

Littlewood, W., & Yu, B. (2011). First language and target language in the foreign language classroom. *Language Teaching, 44*(1), 64–77. https://doi.org/10.1017/S0261444809990310

Liu, X., & Qu, D. (2014). Exploring the multimodality of EFL textbooks for Chinese college students: A comparative study. *RELC Journal, 45*(2), 135–150. https://doi.org/10.1177/0033688214533865

Liu, Y., & O'Halloran, K. L. (2009). Intersemiotic texture: Analyzing cohesive devices between language and images. *Social Semiotics, 19*(4), 367–388. https://doi.org/10.1080/10350330903361059

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

McDonald, P. (2014). Sustainability in CALL learning environments: A systemic functional grammar approach. *The EUROCALL Review, 22*(2), 3–18. https://files.eric.ed.gov/fulltext/ED574679.pdf

Mhlongo, S., Mbatha, K., Ramatsetse, B., & Dlamini, R. (2023). Challenges, opportunities, and prospects of adopting and using smart digital technologies in learning environments: An iterative review. *Heliyon, 9*(6), Article e16348. https://doi.org/10.1016/j.heliyon.2023.e16348

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Multilingual Matters. https://doi.org/10.21832/9781847692092

Misa, M. (2023). Multiliteracies pedagogy: A case study of critical reading in ELT classroom by implementing situated practice. *Journal of English Education, 11*(2), 489–500. https://journal.uniku.ac.id/index.php/ERJEE/article/view/7918/3708

Morell, T., Fernandez-Pacheco, N., Natalia, N., & Beltran-Palanques, V. (2020). How do trained English-medium instruction (EMI) lecturers combine multimodal ensembles to engage their students? In R. Roig-Vila (Ed.), *La docencia en la Enseñanza Superior. Nuevas aportaciones desde la investigación e innovación educativas* (pp. 308–321). Octaedro. https://rua.ua.es/depace/bitstream/10045/110190/1/La-docencia-en-la-Ense nanza- Superior_31.pdf

Motteram, G. (2011). Developing language-learning materials with technology. In B. Tomlinson (Ed.), *Materials development of language teaching* (2nd ed.) (pp. 303–327). Bloomsbury.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–93. https://doi.org/10.17763/haer.66.1.17370 n67v22j160u

O'Halloran, K. L. (2005). *Mathematical discourse: Language, symbolism and visual images*. Continuum.

O'Halloran, K. L., Podlasov, A., Chua, A., & K. L. E, M. (2012). Interactive software for multimodal analysis. *Visual Communication, 11*(3), 363–381. https://doi.org/10.1177/1470357212446414

O'Halloran, K. L., Tan, S., & Smith, B. A. (2016). Multimodal approaches to English for academic purposes. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 256–269). Routledge.

Pan, Y.-C., & Pan. Y.-C. (2010). The use of L1 in the foreign language classroom. *Colombian Applied Linguistics Journal, 12*(2), 87–96. https://doi.org/10.14483/22487085.85

Paskins, Z., Sanders, T., Croft, P. R., & Hassell, A. B. (2017). Exploring the added value of video-stimulated recall in researching the primary care doctor–patient consultation: A process evaluation. *International Journal of Qualitative Methods, 16*(1), Article 719623. https://doi.or g/10.1177/1609406917719623

Pertama, T., Rukmini, D., & Bharati, D. (2018). Implementation of three metafunctions in verbal language and visual image of students' textbook. *English Education Journal, 8*(4), 418–431. https://doi.org/10.15294/eej.v8i4.2532

Plomp, T. (2013). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research* (pp. 9–35). SLO. https://ris.utwente.nl/ws/portalfiles/portal/14472302/Introduction_20to_20education_20design_20research.pdf

Royce, T. (2002). Multimodality in the TESOL classroom: Exploring visual-verbal synergy. *TESOL Quarterly, 36*(2), 191–205. https://doi.org/10.2307/3588330

Royce, T. (2007). Intersemiotic complementarity: A framework for multimodal discourse analysis. In T. D. Royce & W. L. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 63–109). Lawrence Earlbaum Associates, Inc.

Salbego, N., Heberle, V., & Balen, M. (2015). A visual analysis of English textbooks: Multimodal scaffolded learning. *Calidoscópio, 13*(1), 5–13. https://doi.org/10.4013/cld.2015.131.01

Saldana, J. (2014). Coding and analysis strategies. In P. Leavy (Eds.), *The Oxford handbook of qualitative research* (pp. 581–605). Oxford University Press.

Sanchez-Stockhammer, C. (2021). Multimodal cohesion through word formation: Sublexical cohesive ties in online illustrated step-by-step cooking recipes. *Discourse, Context & Media, 43*, Article 100536. https://doi.org/10.1016/j.dcm.2021.100536

Schubert, C. (2021). Multimodal cohesion in persuasive discourse: A case study of televised campaign advertisements in the 2020 US presidential election. *Discourse, Context & Media, 43*, Article 100537. https://doi.org/10.1016/j.dcm.2021.100537

Shih, Y. C. (2014). Communication strategies in a multimodal virtual communication context. *System, 42*, 34–47. https://doi.org/10.1016/j.system.2013.10.016

Sidabutar, U., Sinaga, N. T., Sitorus, N., & Lestari, F. D. (2021). A multimodal analysis on a Vocational English book and its effect on students' English proficiency. *Linguistics and Culture Review, 5*(S3), 1651–1665. https://doi.org/10.21744/lingcure.v5nS3.1951

Sueyoshi, A., & Hardison, D. M. (2005). The role of gestures and facial cues in second language listening comprehension. *Language Learning, 55*(4), 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x

Swain, M., & Lapkin, S. (2000). Task-based second language learning: The uses of the first language. *Language Teaching Research, 4*(3), 251–274. https://doi.org/10.1177/136216880000400304

Teedaaksornsakul, M. (2024). *Design and development of intersemiotically cohesive teaching materials for English for agriculture* [Unpublished doctoral dissertation, Thammasat University]. Thammasat University.

Thomas, N., Bowen, N. E. J. A., Louwe, S., & Nanni, A. (2023). Performing a balancing act: A trioethnography of "foreign" EMI lecturers in Bangkok. In F. Fang & K. P. Pramod (Eds.), *English-Medium Instruction pedagogies in multilingual universities in Asia* (pp. 138–154). Routledge. https://doi.org/10.4324/9781003173137-12

Tomlinson, B. (2023). *Developing materials for language teaching* (2nd ed.). Bloomsbury.

Trisanti, N., Sukyadi, D., & Suherdi, D. (2022). Digital multimodal composing usage in EFL secondary classroom: A study of in-service EFL teachers' perception. In W. Strielkowski, J. M. Black, S. A. Butterfield, C. Chang, J. Cheng, F. P. Dumanig, R. Al- Mabuk, N. Scheper-Hughes, M. Urban & S. Webb (Eds.), *Advances in social science, education and humanities research* (pp. 375–382). Atlantis Press. https://doi.org/10.29 91/978-2-494069-91-6_59

Turnbull, M., & Arnett, K. (2002). Teachers' uses of the target and first languages in second and foreign language classrooms. *Annual Review of Applied Linguistics, 22*, 204–218. https://doi.org/10.1017/S0267190502000119

Weninger, C. (2021). Multimodality in critical language textbook analysis. *Language, Culture and Curriculum, 34*(2), 133–146. https://doi.org/10.1080/07908318.2020.1797083

Yassine, S. (2014). Multimodal design of EFL textbooks: a social semiotic multimodal approach. *Journal of the Association-Institute for English Language and American Studies, 3*(12), 84–90. https://www.anglisticum.org.mk/index.php/IJLLIS/article/view/ 1523/2031

Yu, J., & Liu, X. (2022). Text first or picture first? Evaluating two modes of multimodal input for EFL vocabulary meaning acquisition. *SAGE Open, 12*(3), 1–11. https://doi.org/10.1177/21582440221119469

Yulini, F. (2022). Multimodal learning material in an English-speaking class in Kampung Baluwarti. *Journal of English Language Teaching and Learning (JETLE), 3*(2), 73–79. https://doi.org/10.18860/jetle.v3i2.15878