Scientific and Social Research, 2025, Volume 7, Issue 1
https://ojs.bbwpublisher.com/index.php/SSR
Online ISSN: 2981-9946
Print ISSN: 2661-4332

A Critical Study on Multimodal Composing Competence of University Students from Hong Kong, Macao, and Taiwan Regions

Lu Jiang[1], Yiping Chen[1]*, Zaibo Long[2]

[1]School of Translation Studies, Jinan University, Zhuhai 519070, China

[2]School of Foreign Languages, Wuhan University of Science and Technology, Wuhan 430081, China

Corresponding author: Yiping Chen, tchenyiping@jnu.edu.cn

Abstract: This study investigates the digital multimodal composing competence of fifty freshmen from Hong Kong, Macao, and Taiwan regions in a Chinese University based on a systemic functional approach to multimodal discourse analysis (SF-MDA) and sociosemiotic ethnography. These students are divided into six groups and are required to make presentations on topics like sports, fashion, festivals, environment, and cities. The collected data include students' digital multimodal texts and interview responses. Findings show that a) students have the awareness of salience, but often misrepresent it; b) students are likely to confuse the relationship and function between the text and image as they often use the image with the least or redundant information to complement the text; c) students are likely to incorporate something irrelevant to the topic; d) students misuse the immersive image as a background due to lack of the concept of context; e) they usually provide excessive language information. This study also puts forward suggestions on the pedagogy of improving these students' multimodal literacies through optimizing the existing teaching method and setting up an instantaneous recall feedback form.

Keywords: Multimodal composing competence; Systemic functional approach to multimodal discourse analysis; Multimodality; Pedagogy; Students from Hong Kong, Macao, and Taiwan regions

Online publication: January 23, 2025

1. Introduction

Nowadays, there emerge an increasing amount of studies on multimodal composing, framed by multiliteracies and sociocultural theories, including exploring the affordances of linguistic and non-linguistic semiotic resources and analyzing the process of multimodal composing by individual or in collaboration, and improving multimodal pedagogy [1–14]. Many studies focused on the multimodal products and explored how language learners orchestrated

different kinds of modes into ensembles to fully express the intention and meaning, among which Shin et al. went further to explore why sixth-grade L2 learners orchestrate different modes in this way by adopting the method of a systemic functional approach to multimodal discourse analysis and principles of sociosemiotic ethnography [15–19]. However, few studies have summarized typical problems and general patterns of university students from Hong Kong, Macao, and Taiwan regions in orchestrating different modes before they receive any systematic multimodal knowledge. These students almost all live in a multilingual environment before receiving higher education in China's mainland. Some of them have lived in Hong Kong, Macao, and Taiwan regions since they were born and were accustomed to the education concept, teaching method, and evaluation system there.

To fill the research gap, this study examines the multimodal composing competence of fifty L2 freshmen from Hong Kong, Macao, and Taiwan regions in China by exploiting all available semiotic resources in multimodal PPT. These students have been divided into groups according to different topics. Each group is required to make a PowerPoint about the given topic and do a multimedia presentation in class. Researchers will critically analyze each group's PowerPoints by systemic functional approach to multimodal discourse analysis (SF-MDA) to test whether multimodal ensembles in each PPT slide and PPT slides as a whole have effectively expressed the ideational function, interpersonal function, and textual function or not. In particular, the intermodal relations between language and image will be analyzed. Besides, there will be a short interview after each presentation, concerned with the reasons for orchestrating multimodal resources in the way they did. Finally, the implications of pedagogy for developing students' multimodal composing competence will be discussed. This study specifically sought to answer the following questions by both quantitative and qualitative methods:

1. What are the typical problems of L2 freshmen from Hong Kong, Macao, and Taiwan regions in multimodal composing?

2. Why do they orchestrate different modes in this way?

2. Critical thinking of intermodal relations between language and image from the perspective of SF-MDA

This study utilizes the systemic functional approach to multimodal discourse analysis (SF-MDA) to examine students' presentation slides. SF-MDA, grounded in social semiotics, leverages systemic functional linguistics (SFL) to analyze how various communication modes interact [20–21]. According to SFL, the metafunction of language includes ideational function, interpersonal function, and textual function [22–23].

The ideational function is comprised of experiential and logical components. The experiential function conveys experiences from both external and internal realms, incorporating entities like people, objects, and events, typically represented through transitivity and voice. The interpersonal function reflects the speaker's identity, status, and attitudes, realized through mood, modality, and appraisal systems. Mood includes declarative, interrogative, and imperative structures, while modality addresses the speaker's judgments about propositions. The textual function organizes information through theme and rheme, where the theme conveys given information and the rheme provides new details [24].

Unsworth identifies three intermodal relationships between language and images: concurrence, complementarity, and connection [25]. Concurrence occurs when language and images elaborate on the same information, presenting redundancy, exposition, instantiation, or homospatiality. Complementarity describes how images can add details that language may omit. Connection encompasses projection and conjunction. Projection

involves quoting and thought representation through intermodal links, while conjunction links time, place, and causation using symbols like arrows and circles. These relationships significantly enhance ideational meaning. Interpersonal meaning is conveyed through characters in images. According to Painter et al., images with a gaze engage viewers actively, while those without primarily convey information [26]. Textual meaning is reflected in layout, with familiar information typically on the left and unfamiliar on the right [27].

The SF-MDA method proves useful in examining how students use various semiotic resources — linguistic, visual, aural, gestural, and spatial — to create meaning. This approach explores the grammar of these resources and their combinations [19, 28–29]. It has been applied to analyze pedagogic discourse and student compositions, highlighting the integration of language, gesture, and spatial arrangements in learning contexts [30–31].

In this study, the researchers examine the combined effectiveness of language and images, positing that this synergy should be more diverse and impactful than the individual effects of either modality alone. Since redundancy between language and images does not enhance the overall information conveyed, it should not be encouraged in multimodal composition. The researchers will critically assess whether students select appropriate images that align with the language to effectively achieve ideational, interpersonal, and textual functions.

3. Research method

3.1. Context and participant

This study examines 50 freshmen from Hong Kong, Macao, and Taiwan regions at a prestigious university that promotes a "facing overseas and facing Hong Kong, Macao, and Taiwan regions" policy. These students, raised in multilingual and multicultural environments, often display active thinking and distinct personalities. They gain admission through various ways, including DSE (Diploma of Secondary Education) in Hong Kong, Macao's joint admission exam, national joint college entrance exam, and high school performance recommendations. DSE English scores range from fair (Level 3) to distinguished (Level 5**), while Macao requires a passing score of 60 out of 100. In the Taiwan Region, the GSAT (General Scholastic Ability Test) identifies top performers at levels 11 and 12.

In this English class of fifty students, results show that only four out of thirteen students from Hong Kong achieved satisfactory DSE scores (**Figure 1**). Of the fourteen students who took the national joint college entrance exam, ten passed (**Figure 2**). In Macao, all recommended students qualified, but only three scored above 80 (**Figure 3**). Additionally, two students from the Taiwan region achieved level 11 on the GSAT, while three passed an upgraded recruitment exam. Overall, 58% of the students in this class passed their English examinations, with 36% attaining relatively high scores. They excel in listening and speaking skills but struggle with reading and writing. Furthermore, these students actively engage in English-related activities both inside and outside the classroom, displaying a greater willingness to express themselves compared to their peers from mainland China.

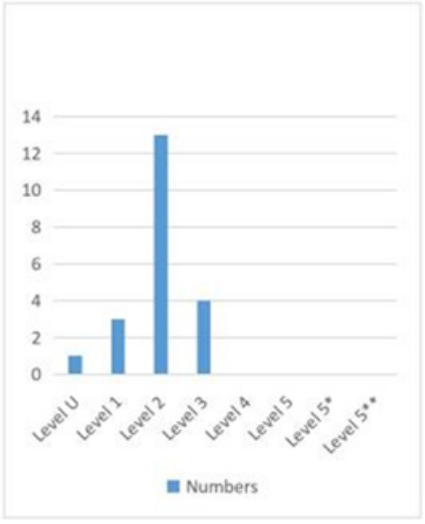

Figure 1. Number of English levels of students from Hong Kong in DSE

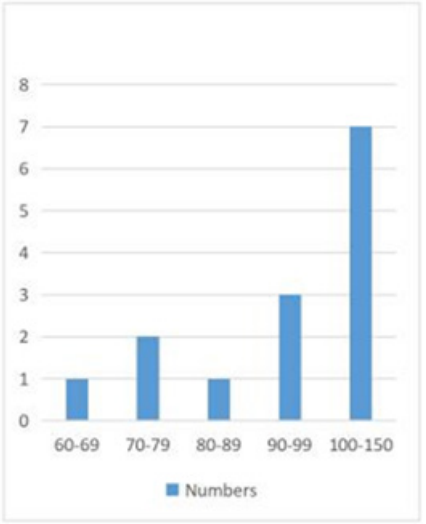

Figure 2. English score of students from Hong Kong, Macao, and Taiwan region in the joint college entrance examination

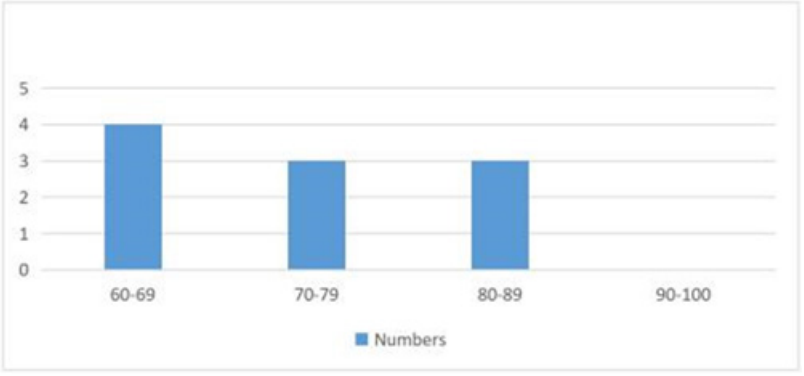

Figure 3. English scores of recommended students from Macao

3.2. Research methods

This study was based on four curricular units in the textbook of New Progressive College English Inspire 1 and New Progressive College English Inspire 2, including topics like sports, fashion, festivals, and environment, and was conducted by both quantitative and qualitative methods. Quantitatively, electronic copies of the PowerPoint slides around four topics were collected and analyzed by the methods of SF-MDA. Qualitatively, prompt interviews were done after each group's presentation. The researcher is their English teacher, and since she did not participate in the student's activity of making PowerPoint slides, nor did she give any suggestions, the interview in a certain sense, can be guaranteed as objective.

4. Findings

This study has summarized students' four major problems of failing to express metafunctions of language and text among 344 PPT slides, including misrepresentation of salience, misuse of immersive pictures, stylistic conflict, little correlation between pictures and texts, and excessive language information.

Table 1 reveals that students generally manage the relationship between images and text well, with only 17.4% of PPT slides failing to express language and text metafunctions. The primary issue identified is "Little correlation between pictures and texts", where students often select irrelevant images to match the text. Additionally, excessive English words or Chinese characters on slides increase reading difficulty, leading audiences to lose interest and overlook essential meanings. Problems of "Misrepresentation of salience" and "Misuse of immersive pictures" are nearly equally represented, as ineffective use of immersive images can downplay critical content. Mistakes related to "Stylistic conflict" are less frequent. It is important to note that multiple issues may occur within a single slide, resulting in a cumulative percentage that does not total 100%. A multimodal analysis expert independently coded the slides, and the comparison of results showed over 85% agreement with the author's findings, ensuring the data's accuracy and validity.

Table 1. Major problems of failing in expressing metafunctions of language and text

Reasons of failing to express metafunctions of language and text	Misrepresentation of Salience	Misuses of immersive pictures	Stylistic conflict	Little correlation between pictures and texts	Excessive language information
Numbers of PPT slides	11	10	4	23	12
Ratio	3.2%	3.0%	1.2%	6.7%	3.5%

4.1. Misrepresentation of salience

According to Kress and van Leeuwen, in spatially integrated texts, salience is assessed through visual cues, such as size, sharpness of focus, tonal and color contrast, placement in the visual field, and perspective [32].

In **Figure 4**, two images are analyzed, a smaller image depicting table tennis equipment and a larger image occupying nearly half the slide, featuring a triangle with trees, rocks, and table tennis items. Despite the larger size of the triangle, the table tennis elements, such as bats, balls, and a net, are minimized and blended with the green trees and gray rocks, making them less noticeable. One student from Group 2 was interviewed to explain the purpose and significance of the trees and rocks in the triangle, as well as the relationship between the rectangular and triangular images.

Researcher: What does this triangle stand for?

Student: It stands for the net, bat, and table tennis.

Researcher: What do you mean by using this triangle?

Student: We want to emphasize the net, table, bat, and ball as we believe that the three angles of the triangle can make them more striking.

Researcher: What's the meaning of trees and rocks in the triangle?

Student: It is meaningless. It has no particular function except to be pleasing to the eyes. Researcher: Does the small picture have the same meaning as the big triangle?

Student: Yes.

Researcher: Why do you choose two pictures with the same meaning? Student: We just want to emphasize the table tennis.

Researcher: Since this topic is about "How to play table tennis", why do you choose these pictures?

Student: Actually, we do not know how to choose the proper picture to illustrate the topic.

The interview reveals that students view the three sharp angles of a triangle as a means of emphasis. It seems plausible because the triangle as a symbol can really play the role of emphasis. However, in this case, the triangle is amplified as a picture with trees, rocks, a net, a bat, and table tennis. Thus, it loses the symbolic function of emphasis. Besides, the net, table, bats, and balls, which should be central, are overshadowed by the green trees and gray rocks. As Tversky et al. and Mayer note, multimedia elements must be relevant to the content to prevent cognitive overload [33–34]. Adding unrelated elements, like trees and rocks, compromises the educational value. Moreover, the student noted that the two pictures, both aiming to emphasize the net, table, bat, and ball, redundantly convey the same information. According to SF-MDA, these images, with little relevance to the topic of "How to play?", fail to effectively convey conceptual meaning. Despite attempts to use the triangle for emphasis and attract attention, irrelevant content undermines both conceptual and interpersonal meaning.

Figure 4. Students' PowerPoint Text on ping pong

4.2. Misuses of immersive pictures

Immersive pictures serve as backgrounds in presentations, providing vivid visual scenes for text. It is a little difficult to use immersive pictures properly in the presentation. As the immersive picture takes up the whole slide, only concise text should be included.

Approximately three percent of immersive pictures in PowerPoint slides are used inappropriately, causing issues such as excessive text, repetitive use leading to aesthetic fatigue, and distraction from the main content when multiple images are present. **Figure 5** from a student presentation on industrial environmental pollution illustrates these problems. While the immersive picture effectively showcases industrial activities, it is overloaded with text, some of which interferes with the image, reducing readability.

Researcher: Why do you use immersive pictures in your presentation?
Student: The immersive picture can provide a vivid scene which can give audiences a visual impact.
Researcher: Why do you use the same immersive picture throughout the whole presentation?
Student: I think it has the same function as a PPT template, for they all can provide a background.

Interviews reveal that students understand the textual and interpersonal meanings of immersive pictures but often confuse them with PowerPoint templates. Templates can be reused without affecting readability as they carry minimal conceptual meaning. In contrast, immersive pictures convey specific meanings, requiring viewers to interpret the relationship between the image and text. Reusing the same immersive picture throughout a presentation hinders readability and leads to aesthetic fatigue. Moreover, a single immersive picture cannot

represent multiple distinct texts, as each relates to a different visual scene.

Figure 5. Students' PowerPoint texts on industrial environmental pollution

4.3. Stylistic conflict

Stylistic conflict refers to the presence of incompatible genres within the same context. In students' presentations, this specifically manifests as the coexistence of contrasting styles on a single PPT slide, such as the juxtaposition of a realistic image with a cartoon, or an incongruous combination of a classical template and modern content.

In **Figure 6**, an immersive image of a lively Mid-Autumn Festival celebration in a square is used as a backdrop, stylistic conflict arises when it is combined with a cartoon image, making the entire slide disjointed. This conflict can occur between images and text or among images themselves. An interview was conducted to explore whether students are aware of these stylistic issues.

Researcher: Have you ever noticed that the immersive picture was taken in the square?

Student: Yes.

Researcher: Why do you insert the cartoon in the immersive picture?

Student: Because we want to highlight the happy and harmonious atmosphere of the Mid-autumn festival with the cartoon.

Researcher: Can't the immersive picture fully convey such meaning?

Student: No. The immersive picture serves as a background and has been fuzzified. Therefore, we insert another picture to highlight the joyous atmosphere.

Researcher: Why do you insert a cartoon rather than a picture of the living scene in real life? Student: But it is more cute, isn't it?

The conversation reveals that students lack an understanding of genre, mistaking the immersive festival image, which features people celebrating with lanterns and humanoid doll lanterns, for a static cartoon. While the doll lanterns may resemble cartoon characters, the image itself portrays dynamic, real-life activities and should not be confused with a cartoon background. According to multimodal discourse analysis, consistency in image genre is crucial for creating a cohesive presentation [35]. Images that align in genre reinforce the message and enhance the narrative, making it easier for the audience to follow. Inconsistent image genres can cause confusion and distract from the main message [36]. Therefore, using images of the same genre as the dynamic, immersive picture is more effective than mixing genres. Consistent genre usage in slides ensures a coherent presentation and leaves a stronger impression on the audience.

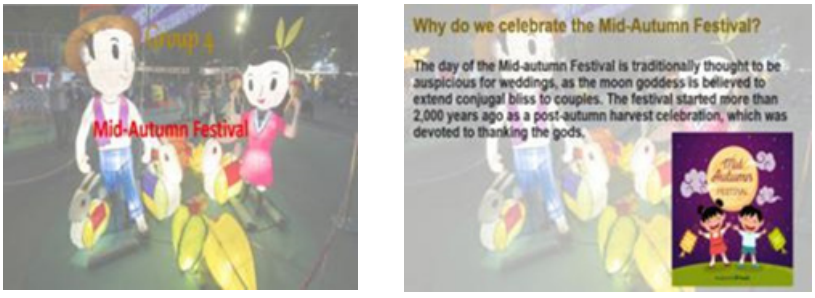

Figure 6. Students' PowerPoint texts on Mid-autumn Festival

4.4. Little correlation between pictures and texts

A weak correlation between images and text suggests a failure of images to elaborate, instantiate, complement, or connect meaningfully with the accompanying text. This disconnect is often due to a misunderstanding of the ideational, interpersonal, and textual meanings of the image.

In **Figure 7**, students presented water pollution using three distinct paragraphs. The first discusses water pollution as a challenge, the second emphasizes water's importance, and the third attributes pollution to industrialization. These paragraphs are not chronologically ordered, yet the accompanying image on the left illustrates a timeline from present to future with arrows and circles representing key events, which do not correspond to the text.

Researcher: Why do you use this picture in the slide? How do you understand this picture?

Student: We think this picture is very simple and clear. Considering that we are going to write an introduction to water pollution with many words, the three arrows in this picture may help the audience understand the meanings clearly as we have three corresponding paragraphs. Each arrow matches with one paragraph. Secondly, the big green circle with the word "NOW" is more than properly fitting our theme because we will talk about the present situation of water pollution.

Researcher: Have you ever thought about the logical meanings of arrows except the function of indexing?

Student: No.

An interview with Group 4's representative reveals a misunderstanding of the logical role of the arrows. Each arrow represents a distinct event on the timeline, which should reflect a progressive rather than parallel relationship. The green and blue arrows serve more than just as indices, as they appear on a dotted-line arrow. Additionally, since each paragraph is already color-coded, using arrows for differentiation is unnecessary.

From a systemic-functional multimodal discourse analysis (SF-MDA) perspective, the diagram does not support the text's conceptual meaning. The timeline in the image suggests events should be sequential, yet the text lacks a chronological structure, undermining the diagram's textual function. This misalignment between text and image results in a failure to effectively convey both ideational and textual meaning, compromising the overall coherence of the presentation.

Figure 7. Students' PowerPoint texts on water pollution

4.5. Excessive language information

Given the limited space on PPT slides, it is important to provide a concise amount of text. Specifically, five or six lines in size 4 font are ideal. Excessive text may crowd out images, negatively impacting the text's effectiveness (**Figure 8**). Additionally, a high word count can quickly lead to reader fatigue, undermining the slide's interpersonal function. Once interest wanes, the overall conceptual meaning may be lost. An interview regarding motivation is presented below.

Figure 8. Students' PowerPoint texts on environmental pollution

Researcher: Have you noticed that there are excessive words in the PowerPoint Texts?

Student: Yes.

Researcher: Why do you put so many words on the slide?

Student: Because I am afraid that my English is poor and I cannot remember all the information. If I put them all on the slide, I would not feel anxious when I do the presentation.

Researcher: But it will increase the reading burden of readers and make them lose interest instantly.

Student: I know.

The conversation above illustrates that while students recognize the drawbacks of overcrowding slides with text, they prioritize their own sense of security during presentations. They fear the embarrassment of forgetting their words, making mistakes, and losing face in front of an audience. To address this, it is essential to provide students with more encouragement and support, helping them to rely less on text during their presentations and fostering their confidence.

5. Pedagogical implications

Past research emphasized the challenge for teachers and the importance of supporting students' digital composing practices [13, 28, 37–38]. Unsworth and Mills proposed a practical pedagogy that incorporates non-linear planning, explicit grammatical design, and playful interaction [30]. Liang and Lim and Lim and Unsworth piloted a pedagogical framework that includes teaching metalanguage, guiding the design thinking process, introducing digital tools and methods, and providing structured lesson packages [14, 39]. These pedagogies generally follow a process-oriented approach rather than targeting specific issues, emphasizing the orchestration of language and video. However, the ability to integrate language and images effectively into texts is fundamental for creating compelling digital videos. Without strong multimodal orchestration skills, students struggle to produce quality content. Moreover, while Lim et al. and Lim et al. demonstrated that teaching metalanguage can enhance multimodal literacy, many students find it challenging to grasp these concepts without familiarity with the relevant terminologies [40–41]. Therefore, it is essential to optimize teaching methods to alleviate students' anxiety and frustration from the outset.

This paper aims to optimize teaching through representative examples. Educators can teach multimodality systematically by analyzing common mistakes in students' presentations. By comparing slides before and after modification, students can reflect on multimodal design and text organization. In each course, only one or two modes should be explicitly taught, and students must apply these concepts to specific contexts. For instance, **Figure 9** from Group 3's presentation on green energy illustrates a significant disconnect with the text which discusses the joint efforts needed from the government and citizens in Hong Kong to promote green energy. However, the image, a cold desktop computer display, fails to convey either ideational or interpersonal meaning. This PPT slide serves as an example for teaching the relationship between text and image by analyzing its shortcomings and proposing an alternative visual. Students can be guided to explore the differences in ideational, interpersonal, and textual meanings between the two images. **Figure 10** effectively represents a positive vision of green energy in Hong Kong, complementing the text appropriately.

Figure 9. Students' PowerPoint texts on city (Hong Kong) problems and solutions

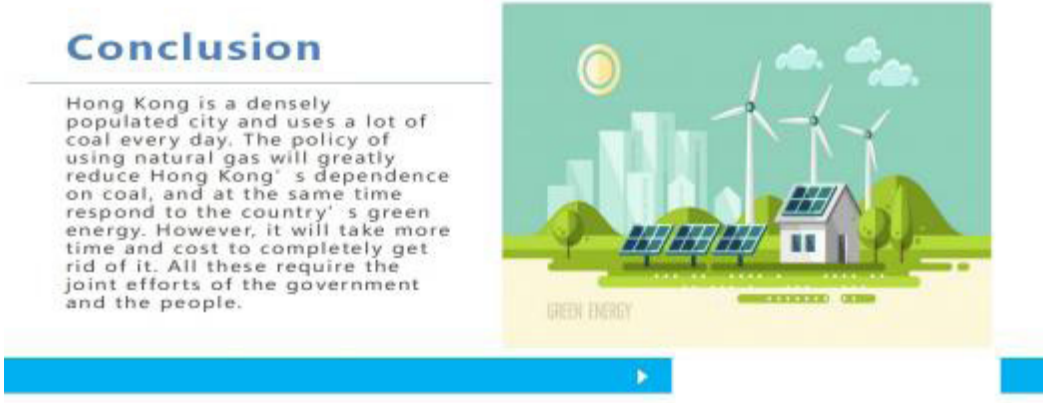

Figure 10. Revised PowerPoint texts on city (Hong Kong) problems and solutions

After class, groups should exchange their PPT slides. Team members must read, select, and comment on slides that align with the multimodal knowledge, such as the appropriate use of immersive images and effective presentation of salient elements. They should return the slides to the original group with their feedback. If disagreements arise, teachers should encourage students to discuss these differences before addressing them in class. This approach not only provides effective training during lessons but also reinforces multimodal knowledge in students' independent work.

In addition to enhancing students' abilities to identify relationships between texts and images, it is essential to cultivate their assessment skills. Hafner and Ho advocated for diversified assessments of multimodal composing [42]. Previous studies have supported the idea that peer review serves as an effective formative assessment method for learning multimodal concepts [43–44].

Recent studies have focused on assessing digital multimodal composing competence [45–48]. Various rating scales have been developed and refined, addressing linguistic, visual, gestural, auditory, and spatial modes as well as criteria related to context, substance, and organization [43, 45]. However, current assessment tools are primarily designed for video products. PPT presentations emphasize the relationship between text and images and the

speaker's language relative to slide content, whereas video assessments focus on factors such as facial expressions, actions, and cinematography. Thus, these criteria may not adequately evaluate multimodal literacy specific to PPT presentations.

The multimodal analysis of PPT slides is as crucial as analyzing digital video, particularly due to the nature of presentations. PPT slides synthesize complex information concisely, requiring effective organization and spatial deployment. Additionally, PPT slides are commonly used in academic and business contexts, making their effective design vital for communicative efficiency.

To effectively enhance students' multimodal literacy in PPT presentations, this paper proposes an instantaneous recall feedback form (**Table 2**) with two peer review questions: "What inspires or impresses you most in the other group's presentation?" and "What confuses you most regarding the text and image?" Students are required to answer these questions and provide justifications immediately after the presentation. This feedback mechanism assesses students' instantaneous comprehension and helps presenters recognize which modes were effectively utilized, which materials were selected, and which conceptual frameworks and language expressions effectively engaged their peers.

Table 2. Instantaneous recall feedback form

Questions	Answers	Reasons
What inspires or impresses you most in other group's presentations in terms of the text and picture?		
What confuses you most in other group's presentation in terms of the text and picture?		

6. Conclusion

Previous studies have discussed some major problems of PPT slides made by students, such as modality overload, misalignment with learning objectives, and poor visual design. Students often overload a single slide with excessive information and text, resulting in a monotonous reading experience instead of an engaging presentation [49–50]. Slides that deviate from the main topic or contain irrelevant information confuse the audience and diminish educational value [51]. Additionally, inadequate color use, inappropriate image selection, and lack of visual hierarchy make slides challenging to interpret [50]. This study confirms that students from Hong Kong, Taiwan, and Macao Regions at a Chinese university encounter similar issues when creating PPT slides, but also face new challenges, such as misrepresentation of salience, misuse of immersive images, and stylistic conflict. Moreover, previous studies primarily described these problems, rarely addressing their underlying causes. Understanding why students make such mistakes in PPT design is crucial for enhancing their multimodal composing skills.

This study examined the digital multimodal composing abilities of fifty freshmen from Hong Kong, Macao, and Taiwan regions using a systemic functional approach to multimodal discourse analysis (SF-MDA) and sociosemiotic ethnographic methods, before any formal multimodal instruction. Findings reveal five major issues affecting the expression of metafunctions in language and text: misrepresentation of salience, misuse of immersive images, stylistic conflict, poor correlation between images and text, and excessive language information. In today's digital era, multimodal composing is an essential literacy. Analyzing these students' abilities can not only help them express themselves more effectively on social media and excel in their future careers but also assist teachers in designing targeted and purposeful instruction grounded in systematic multimodal knowledge.

This study also offers pedagogical implications for enhancing students' multimodal literacy, including

optimizing teaching methods through case studies and implementing immediate feedback mechanisms. However, multimodal composing encompasses more than just the integration of semiotic resources such as language and images. It also involves the presentation and communication of language alongside other modes like sound, gesture, expression, and video. Further exploration is needed to understand how students create digital video productions and cohesively utilize linguistic and non-linguistic resources.

Funding

This research was funded by Guangdong Province Philosophy and Social Science Planning 2023 Annual Joint Research Project (Approval Number: GD23XXW15) and Jinan University 23rd Batch of Teaching Reform Research Project.

Disclosure statement

The authors declare no conflict of interest.

References

[1] New London Group, 1996, A Pedagogy of Multiliteracies: Designing Social Features. Harvard Educational Review, 66(1): 60–92.

[2] Street B, 1984, Literacy in Theory and Practice. Cambridge University Press, Cambridge.

[3] Dalton BK, Robinson J, Lavvorn BE, et al., 2015, Fifth-grade Students' Multimodal Compositions: Modal Use and Design Intentionality. Elementary School Journal, 115(4): 548–569.

[4] Takayoshi P, Seife CL, 2007, Thinking about Multimodality, in Multimodality Composition: Resources for Teachers. Hampton, New Jersey, 1–12.

[5] Vasudevan L, Schultz K, Bateman J, 2010, Rethinking Composing in a Digital Age: Authoring Literate Identities through Multimodal Storytelling. Written Communication, 27(4): 442–468.

[6] Beach R, O'Brien D, 2015, Fostering Students' Science Inquiry through App Affordances of Multimodality, Collaboration, Interactivity, and Connectivity. Reading & Writing Quarterly, 31(2): 119–134.

[7] Burnett C, 2016, Being Together in Classrooms at the Interface of the Physical and Virtual: Implications for Collaboration in On/off-screen Sites. Learning, Media and Technology, 41(4): 566–589.

[8] Smith BE, 2019, Collaborative Multimodal Composing: Tracing the Unique Partnerships of Three Pairs of Adolescents Composing across Three Digital Projects. Literacy, 53(1): 14–21.

[9] Ito M, Horst H, Bittanti M, et al., 2010, Hanging Out, Messing around and Geeking out: Kids Learning and Living in New Media. MIT Press, Cambridge.

[10] Jocius RK, 2018, Becoming Entangled: An Analysis of 5th Grade Students' Collaborative Multimodal Composing Practices. Computers and Composition, 2018(47): 14–30.

[11] Jiang LJ, Hafner C, 2024, Digital Multimodal Composing in L2 Classrooms: A Research Agenda. Language Teaching, 1–19. https://doi.org/10.1017/S0261444824000107

[12] Kim Y, Kang S, Nam Y, et al., 2022, Peer Interaction, Writing Proficiency, and the Quality of Collaborative Digital Multimodal Composing Task: Comparing Guided and Unguided Planning. System, 106(4): 102722. https://doi.org/10.1016/j.system.2022.102722

[13] Chandler PD, 2017, To What Extent Are Teachers Well Prepared to Teach Multimodal Authoring? Cogent Education, 4(1): 1–19.

[14] Liang WJ, Lim FV, 2020, A Pedagogical Framework for Digital Multimodal Composing in the English Language Classroom. Innovation in Language Learning and Teaching, 15(4): 306–320. https://doi.org/10.1080/17501229.2020.1800709

[15] Ajayi L, 2009, English as a Second Language Learners' Exploration of Multimodal Texts in a Junior High School. Journal of Adolescent & Adult Literacy, 2009(52): 585–595.

[16] Toohey K, Dagenais D, Fodor A, et al., 2015, That Sounds so Cooool: Entanglements of Children, Digital Tools, and Literacy Practices. TESOL Quarterly, 49(3): 461–485.

[17] Shin D, 2018, Multimodal Mediation and Argumentative Writing: A Case Study of a Multilingual Writer's Metalanguage Awareness Development, in Bilingual Learners and Social Equity: Critical Approaches to Systemic Functional Linguistics. Springer, New York, 225–242.

[18] Smith B, Pacheco M, de Almeida CR, 2017, Multimodal Codemeshing: Bilingual Adolescents' Processes Composing across Modes and Languages. Journal of Second Language Writing, 2017(36): 6–22.

[19] Shin D, Cimasko T, Yi Y, 2020, Development of Metalanguage for Multimodal Composing: A Case Study of an L2 Writer's Design of Multimedia Texts. Journal of Second Language Writing, 2020(47): 100714. https://doi.org/10.1016/j.jslw.2020.100714

[20] Kress G, van Leeuwen T, 2001, Multimodal Discourse: The Modes and Media of Contemporary Communication. Oxford University Press, New York.

[21] Prior P, 2005, Moving Multimodality beyond the Binaries: A Response to Gunther Kress' "Gains and Losses". Computers and Composition, 2005(22): 23–30.

[22] Halliday MAK, 1985, Spoken and Written Language. Deakin University Press, Geelong.

[23] Halliday MAK, Matthiessen CMIM, 2014, An Introduction to Functional Grammar. Hodder Arnold, London.

[24] Eggins S, 2004, An Introduction to Systemic Functional Linguistics. Bloomsbury, London.

[25] Unsworth L, 2006, E-literature for Children: Enhancing Digital Literacy Learning. Routledge, England.

[26] Painter C, Martin JR, Unsworth L, 2013, Reading Visual Narratives: Image Analysis of Children's Picture Books. Equinox, London.

[27] Kress G, van Leeuwen T, 2006, Reading Images: The Grammar of Visual Design. Routledge, London.

[28] Unsworth L, 2006, Towards a Metalanguage for Multiliteracies Education: Describing the Meaning-making Resources of Language-image Interactions. English Teaching: Practice & Critique, 5(1): 55–76.

[29] Jewitt C, Bezemer J, O'Halloran K, 2016, Introducing Multimodality. Routledge, London.

[30] Unsworth L, Mills KA, 2021, The Affordances of Multimodal Literacy Practices in Digital and Non-digital Learning Contexts for Language and Literacy Learners. Literacy, 55(2): 75–85. https://doi.org/10.1111/lit.12212

[31] Shanks N, Bakay L, 2022, Examining the Impact of Multimodal Composing on English Learners' Academic Literacy Skills. Journal of Adolescent & Adult Literacy, 65(6): 563–573. https://doi.org/10.1002/jaal.1167

[32] Kress G, Van Leeuwen T, 2020, Reading Images: The Grammar of Visual Design. Routledge, London.

[33] Tversky B, Morrison JB, Betrancourt M, 2002, Animation: Can it Facilitate? International Journal of Human-Computer Studies, 57(4): 243–253.

[34] Mayer RE, 2001, Multimedia Learning. Cambridge University Press. Cambridge.

[35] Pirini J, 2017, Approaches to Multimodal Discourse Analysis. Oxford Bibliographies in Communication. https://doi.org/10.1093/obo/9780199756841-0183

[36] Wilmes F, Siry C, 2020, Multimodal Interaction Analysis: a Powerful Tool for Examining Plurilingual Students' Engagement in Science Practices. Journal of Science Education and Technology, 29(6): 1–15. https://doi.org/10.1007/s11165-020-09977-z

[37] Hansford D, Adlington R, 2009, Digital Spaces and Young People's Online Authoring: Challenges for Teachers. Australian Journal of Language and Literacy, 32 (1–2): 55–68.

[38] Rose D, Martin J, 2012, Learning to Write, Reading to Learn: Genre, Knowledge and Pedagogy of the Sydney School. Equinox, Oakville.

[39] Lim FV, Unsworth L, 2023, Multimodal Composing in the English Classroom: Recontextualising the Curriculum to Learning. English in Education, 57(2): 102–119. https://doi.org/10.1080/04250494.2023.2187696

[40] Lim FV, Tan-Chia L, 2018, Designing Learning for Multimodal Literacy: Teaching Viewing and Representing. Routledge, London.

[41] Lim FV, Cope B, Kalantzis M, 2022, A Pedagogic Metalanguage for Learning: Rebalancing the Cognitive with the Socio-Material. Frontiers in Communication, 2022(7): 830613. https://doi.org/10.3389/fcomm.2022.830613

[42] Hafner CA, Ho WYJ, 2020, Assessing Digital Multimodal Composing in Second Language Writing: Towards a Process-based Model. Journal of Second Language Writing, 2020(47): 100710.

[43] Hung HT, Chiu YC J, Yeh HC, 2013, Multimodal Assessment of and for Learning: A Theory-driven Design Rubric. British Journal of Educational Technology, 44(3): 400–409. https://doi.org/10.1111/j.1467-8535.2012.01337.x

[44] Alt D, Raichel N, 2022, Problem-Based Learning, Self- and Peer Assessment in Higher Education: Towards Advancing Lifelong Learning Skills. Research Papers in Education, 37(3): 370–394. https://doi.org/10.1080/02671522.2020.1849371

[45] Dzekoe R, 2017, Computer-based Multimodal Composing Activities, Self-revision, and L2 Acquisition through Writing. Language Learning & Technology, 21(2): 73–95. https://doi.org/10125/44612

[46] Lee SY, Lo YH, Chin TC, 2019, Practicing Multiliteracies to Enhance EFL Learners' Meaning-making Process and Language Development: A Multimodal Problem-based Approach. Computer Assisted Language Learning, 34(1–2): 66–91. https://doi.org/10.1080/09588221.2019.1614959

[47] Li M, 2020, Multimodal Pedagogy in TESOL Teacher Education: Students' Perspectives. System, 2020(94): 102337. https://doi.org/10.1016/j.system.2020.102337

[48] Zhang ED, Yu S, 2023, Implementing Digital Multimodal Composing in L2 Writing Instruction: A Focus on Developing L2 Student Writers. Innovation in Language Learning and Teaching, 17(4): 769–777.

[49] Zheng YL, 2019, Meaning Construction of Multimodal Discourse in High School English PPT Courseware—A Case Study of the 10th National Demonstration Lectures for High School English Teachers. Theory and Practice in Language Studies, 9(9): 1134–1139. http://dx.doi.org/10.17507/tpls.0909.09

[50] Ledin P, Machin D, 2019, Doing Critical Discourse Studies with Multimodality: From Metafunctions to Materiality. Critical Discourse Studies, 16(5): 497–513. https://doi.org/10.1080/17405904.2018.1468789

[51] Feng D, 2021, Genre, Pedagogy, and PowerPoint Design: A Multimodal Move Analysis of Linguistics Lecture Slideshows, in Approaches to Specialized Genres: In Memory of Stephen Evans. Routledge, London.