Studies in Second Language Learning and Teaching
Department of English Studies, Faculty of Pedagogy and Fine Arts, Adam Mickiewicz University, Kalisz
SSLLT 0 (0). 2026. 1-30. Published online: 07.09.2026
https://doi.org/10.14746/ssllt.53401
http://pressto.amu.edu.pl/index.php/ssllt

Writing quality, vocabulary learning, and reading comprehension in integrated reading-to-write tasks: A comparison of digital multimodal composing and traditional monomodal writing

HeeJoon Choi
Georgia State University, Atlanta, USA
https://orcid.org/0009-0007-4859-6930
hchoi49@gsu.edu

Yingliang Liu
South-Central Minzu University, Wuhan, China
https://orcid.org/0000-0002-0538-2752
yliu@mail.scuec.edu.cn

YouJin Kim ✉
Georgia State University, Atlanta, USA
https://orcid.org/0000-0002-5741-9369
ykim39@gsu.edu

Abstract

Integrated reading-to-write tasks have been widely used to enhance second language (L2) learners' writing skills and benefit other areas of L2 learning, such as reading and vocabulary. Although these benefits of integrated tasks have been widely investigated in traditional text-based monomodal writing, few studies have examined pedagogical interventions involving digital multimodal composing (DMC). This study investigates whether and to what extent writing modalities (DMC vs. traditional monomodal writing) in integrated reading-to-write tasks impact writing quality, vocabulary learning, and reading comprehension.

Thirty-six English as a foreign language students in China completed either an argumentative monomodal essay or a DMC video-making task, both of which required incorporating key information from an article and presenting their own argument. Students' writing was assessed using an analytic rubric. A posttest was conducted to assess their reading comprehension (via multiple-choice questions and summary writing) and vocabulary knowledge of words from the source text. Results indicated that the DMC group achieved higher scores in organization (writing quality) and in summary writing (reading comprehension). However, no significant differences were found in other writing scores (content, language use), vocabulary learning, or multiple-choice comprehension.

Keywords: integrated writing; digital multimodal composing; writing performance; reading comprehension; vocabulary learning

1. Introduction

Integrated tasks have been a popular pedagogical practice in second language (L2) instruction, as they combine receptive skills (e.g., reading, listening) with productive skills (e.g., writing, speaking) and thus can foster learners' overall L2 development (Hirvela, 2016). Prior studies in second language acquisition (SLA) and literacy education have explored the benefits of integrated instruction across multiple domains of L2 learning, such as the relationship between reading and writing performance (e.g., Esmaeili, 2002; Shin & Ewert, 2015), writing processes and source comprehension (e.g., Payant et al., 2019), and vocabulary learning (e.g., Jung, 2020). However, much of this research has focused on traditional monomodal compositions, with fewer studies examining how different task modalities in integrated writing might influence L2 learning.

With growing interest in the role of technology in language instruction, L2 writing research has increasingly incorporated digital tools. Digital multimodal composing (DMC), a textual practice involving multiple semiotic resources such as text, audio, and images, has been proposed to expand the role of writing as a means of communication. While previous studies have demonstrated the potential of DMC to enhance L2 writing (e.g., Kim & Belcher, 2020), its effects on other dimensions of L2 learning remain underexplored (Kessler, 2024). Examining its impact on the development of foundational L2 skills, such as vocabulary knowledge and reading comprehension, is therefore important, especially considering concerns (e.g., Qu, 2017) that the cognitive demands of multimodal composing may distract learners from L2 learning. This gap is particularly evident in the context of integrated reading-to-write tasks (Cho & Kim, 2024), where limited research has investigated how DMC may impact L2 outcomes.

To address this gap, the current study investigates how composition modality (DMC vs. traditional monomodal writing) affects English as a foreign language (EFL) learners' writing quality, vocabulary learning, and reading comprehension in reading-to-write tasks. In doing so, it explores the potential of DMC not only as a writing practice but also as a pedagogically meaningful approach to supporting diverse aspects of L2 development.

2. Literature review

DMC is "a textual practice that involves the use of digital tools to produce texts by combining multiple semiotic modes" (Jiang, 2017, p. 413). The theoretical foundation of DMC lies in the concept of multiliteracies proposed by the New London Group (1996). Multiliteracies pedagogy reflects an evolving understanding of language learning in the current digital world. Specifically, multiliteracies pedagogy challenges the traditional view of literacy pedagogy that is "restricted to formalized, monolingual, monocultural, and rule-governed forms of language" (p. 61). Instead, it views language and other modes of meaning as "dynamic representational resources, constantly being remade by their users" (p. 64). In practice, DMC allows learners to create multimodal texts in various formats, including videos, posters, slide shows, or podcasts, which are common forms of digital communication (Jiang & Hafner, 2025). Unlike traditional print-based writing, which primarily relies on linguistic encoding, DMC requires learners to engage in complex multimodal orchestration. This means L2 writers must coordinate visual, auditory, and spatial elements along with text, making decisions about how semiotic resources interact to achieve communicative goals (Kress, 2010; New London Group, 1996).

Despite some concerns that the use of multiple modalities may hinder L2 learning (Casanave, 2017; Qu, 2017), studies increasingly highlight the positive impact of DMC, such as enhanced motivation, autonomy, and identity (e.g., Hafner & Miller, 2011; Jiang & Luk, 2016). However, the effects of DMC on learning remain relatively underexamined (Jiang & Hafner, 2025; Kessler, 2024), and much of the existing work has focused primarily on writing quality and development, leaving other aspects of L2 learning unexplored. More studies are therefore needed to examine how DMC can support L2 learning as a pedagogical practice. The following sections review research comparing DMC and traditional monomodal writing, the relationships between integrated writing and reading comprehension, and how multimodality can support L2 vocabulary learning.

2.1. Comparing DMC and traditional monomodal writing

Recent studies comparing DMC and traditional monomodal writing have increasingly highlighted that DMC can support L2 writing development. For example, Xu (2023) compared Chinese EFL students' writing performance between DMC (collaborative) and traditional (individual) writing conditions and found that the DMC group outperformed the traditional group on the posttest with respect to five aspects of writing quality: content, comprehensibility, mean length per T-unit, task fulfillment, and text length. Other studies have reported no significant differences between DMC and traditional writing conditions. Kim and Belcher's (2020) study explored Korean undergraduate EFL students' linguistic performance by having the same learners complete both a traditional argumentative essay and a DMC project on the same topic. The results suggested that the traditional essays were more syntactically complex, but there were no significant differences in accuracy. Similarly, Kim et al. (2023), employing a between-groups design over a semester in a university classroom in Korea, reported that the students in the two groups showed comparable levels of linguistic complexity and accuracy in their written outputs.

In the context of integrated reading-to-write tasks, Cho (2022) explored Korean university students' writing performance in a book summary-reflection task in an extensive reading context. Participants were randomly assigned to complete a book review task using either DMC or traditional monomodal writing. The results showed no significant differences in linguistic outcomes between the DMC and traditional groups. Similarly, Cho and Kim (2024) conducted a within-participant comparison study of Korean EFL high school students in a storybook summary-reflection task and found no significant difference in writing outcomes. Although neither study showed that one modality clearly outperformed the other, they suggest that engaging in DMC does not necessarily distract learners from linguistic development and that both DMC and traditional writing may lead to comparable benefits in L2 writing development.

Building on evidence that DMC can be as linguistically beneficial as traditional writing, additional evidence is needed on how DMC supports L2 development in integrated tasks. Moreover, while much of the existing research has focused on writing, little attention has been paid to reading or vocabulary development through DMC.

In terms of research design, many previous DMC comparison studies have adopted within-participant designs rather than between-group comparisons (e.g., Cho & Kim, 2024) or have focused on collaborative DMC (e.g., Xu, 2023). Some collaborative DMC studies highlight how peer interaction provides valuable opportunities for collective scaffolding and language development (Akoto & Li, 2025; Nishioka, 2016). Yet, Li and Pham (2025) compared the quality of DMC tasks completed in collaborative versus individual conditions among English for

Specific Purposes (ESP) students creating digital infographics at a Vietnamese university, finding no statistically significant differences in the overall rubric-based DMC scores or linguistic accuracy. Furthermore, as it is difficult to identify each student's exact linguistic contribution in collaborative DMC (e.g., Kim et al., 2022; Xu, 2023), using an individual writing design offers a more accurate view of learners' independent L2 development. Therefore, the current study examines an integrated argumentative writing task completed individually.

2.2. Integrated writing and reading comprehension

In contrast to conventional writing that treats reading and writing as separate skills, integrated tasks have been widely recognized for promoting learners' communicative competence (Cumming, 2014) and supporting language use that resembles real-world meaning-making processes. Integrated reading-to-write tasks require learners to utilize multiple skills beyond writing ability, including comprehending the source text, identifying the main ideas, synthesizing, summarizing, paraphrasing, organizing, and connecting ideas from the source text with their own perspectives (Knoch & Sitajalabhorn, 2013). Therefore, successful performance in integrated tasks often depends not only on writing ability but also on other L2 skills.

Additionally, writing and reading are widely recognized as mutually supportive processes in language learning. For example, in studies examining L2 English undergraduate students enrolled in English for Academic Purposes (EAP) and English as a second language (ESL) programs at universities in Canada and the United States, Payant et al. (2019) and Shin and Ewert (2015) found that reading comprehension scores were significant predictors of writing performance in integrated reading-to-write tasks. Similarly, Esmaeili (2002) observed that first-year engineering students in an ESL writing course at two Canadian universities performed significantly better in both writing and reading recall when the two tasks were thematically related. These findings suggest that students with stronger reading comprehension are more likely to succeed in integrated writing tasks, highlighting that reading can predict writing. However, much existing research has emphasized writing as the dependent variable (i.e., examining how reading influences writing outcomes). It remains unclear how reading-to-write tasks, especially through multimodal practices, may also support learners' reading comprehension outcomes. This potential learning effect may arise because, in multimodal tasks, learners engage in resemiotization, a process in which meanings are translated and shifted across different semiotic modes (Cimasko & Shin, 2017). We therefore hypothesize that the cognitive demands of transmediation, which refers to the process of orchestrating multiple modalities such as text, images, and audio

to represent source information, encourage L2 learners to process the source text more deeply and critically than traditional writing would. To test this hypothesis, we compared reading comprehension between DMC and traditional writing conditions.

Building on this idea, only a few studies to date have examined how multimodal writing practices relate to L2 reading comprehension. Among them, Yang and Wu (2012) compared instruction involving digital storytelling (students completing collaborative DMC) with a lecture-type approach (in which the instructor delivered content via presentation software and students completed individual paper-based writing) in terms of L2 learning outcomes among 10th-grade EFL students in high school English classes in Taiwan. To measure learning outcomes, students took an English achievement test developed for a semester-based course assessment, which included writing, reading, and vocabulary sections. The results showed that the DMC group outperformed the traditional group in reading and writing performance, while vocabulary learning did not differ significantly.

Marino (2025) also examined multimodal writing and L2 reading skills. Beginner-level Italian L2 learners at a US university engaged in collaborative writing tasks, and the study compared DMC and traditional writing conditions across several L2 skills, including overall course achievement, grammatical competence, listening, reading, writing, and speaking. L2 performance was measured by a test developed for a semester-based course assessment. For writing specifically, scores were based on an analytic rubric measuring content, vocabulary, language use, and text length. The results indicated that the DMC group outperformed the control group in language use in writing, but no significant differences were found in reading skills. Marino attributed this lack of significant differences to the nature of the tasks, noting that the tasks did not directly require the learners to engage in receptive skills such as reading. However, since previous monomodal integrated reading-to-write studies have reported that students can perform significantly better in both reading and writing when the two tasks are thematically related (e.g., Esmaeili, 2002), the weak thematic relationship between the reading test (final exam) and course writing tasks in Marino's (2025) study may not have been sufficient to capture any potential reading comprehension gains derived from the DMC tasks.

Furthermore, while a few recent studies have begun to explore integrated reading-to-write tasks, particularly within DMC contexts (e.g., Cho, 2022; Cho & Kim, 2024), their primary focus has been writing outcomes. Cho (2022), for example, found that L2 reading proficiency did not significantly predict the quality of DMC task performance. Although these studies provide empirical evidence for the benefits of DMC in L2 writing, its direct impact on reading comprehension, particularly in argumentative writing contexts, remains underexplored. Addressing these gaps, the current study focuses on individual writing performance using task-specific measures to better understand the impact of task modality.

2.3. Multimodality and L2 vocabulary learning

In addition to reading skills, vocabulary knowledge plays a crucial role in L2 reading-to-write task outcomes. The potential of multimodality for vocabulary acquisition is supported by cognitive theories. According to Paivio's (1986) dual coding theory, the human mind possesses separate but interconnected information processing systems for verbal and nonverbal (i.e., imagery) representations, and memory is maximized when learners build referential connections between verbal and nonverbal processing. Building upon this, Mayer's (2001) cognitive theory of multimedia learning posits that meaningful learning occurs when learners actively engage in selecting, organizing, and integrating both verbal and pictorial models. Because DMC requires learners to actively select, organize, and integrate verbal and nonverbal representations, it helps them build strong referential connections across different processing channels. Furthermore, Craik and Lockhart's (1972) levels of processing framework suggests that deeper processing leads to better retention. In the context of DMC, the act of orchestrating multiple semiotic resources to represent meaning can lead to high levels of cognitive engagement, promoting deeper processing of lexical items.

Building on these claims, previous studies have demonstrated that multimodal input and tools facilitate deeper vocabulary processing and lexical richness across diverse instructional contexts (e.g., Ajayi, 2008; Lee et al., 2021). Although these studies support the role of multimodal input in vocabulary acquisition, only a few studies have explored the potential of multimodal writing tasks in vocabulary learning. Vandommele et al. (2017) examined beginner-level Dutch learners' multimodal writing tasks. Their findings showed that task-based multimodal writing interventions significantly enhanced L2 learners' lexical diversity and text length. More recently, Mai et al. (2023) investigated how multimodality supports vocabulary acquisition. In their study, 63 EFL university students were randomly assigned to two groups and learned vocabulary by either receiving multimodal input (MI group) or creating DMC work (DMC group). The results showed that the DMC group outperformed the MI group on both the posttest and delayed posttest. More relevant to the context of integrated reading-to-write tasks, Cho (2022) compared incidental vocabulary learning between DMC and traditional monomodal writing groups after an extensive reading task. The results revealed that both groups successfully acquired target vocabulary with no significant difference between the modalities, indicating that DMC is at least as effective as traditional writing for vocabulary development. These findings contribute to our understanding of multimodal writing as a potential pedagogical tool for learning vocabulary. However, research on its use in integrated tasks, particularly in argumentative writing contexts, remains limited, warranting further investigation.

2.4. The current study

Taken together, the findings of the extant research reported in the previous sections point to the potential of integrated tasks not only to enhance writing quality, but also to foster vocabulary growth and reading comprehension, particularly when accompanied by multimodal elements. However, empirical evidence remains limited regarding how DMC, as compared to traditional monomodal writing, can support different areas of L2 learning within an integrated task context. Accordingly, the current study explores how composition modality (DMC vs. traditional monomodal writing) affects Chinese undergraduate EFL students' L2 learning in integrated tasks, with a focus on three areas of L2 learning: writing quality, vocabulary learning, and reading comprehension. In addition, while some previous empirical studies have compared DMC and traditional writing, they generally did not control for learners' baseline writing proficiency, despite recent evidence that L2 writing proficiency is a significant predictor of DMC outcomes (Cho, 2022; Cho & Kim, 2024; Kim et al., 2022). To address these gaps, the current study explicitly controls for writing proficiency as a covariate and is guided by the following research questions:

RQ1: Are there any differences in Chinese EFL students' writing quality on an integrated reading-to-write task between the DMC and traditional writing conditions, while controlling for writing proficiency?
RQ2: Are there any differences in Chinese EFL students' learning of source text vocabulary between the DMC and traditional writing conditions, while controlling for writing proficiency?
RQ3: Are there any differences in Chinese EFL students' reading comprehension of the source text between the DMC and traditional writing conditions, while controlling for writing proficiency?

3. Methods

3.1. Research context and participants

This quasi-experimental study was conducted in an EFL context at a Chinese university. The participants were L2 English learners majoring in English and enrolled in a two-credit academic writing course titled *English Writing A2,* which met once a week for 90 minutes. The overall objective of the English Writing A2 course was to develop students' argumentative and research writing skills. As part of the course, students read articles on critical topics in contemporary society, such as multiculturalism, gender inequality, and the purpose of education, and engaged

in integrated writing tasks. Throughout the semester, students completed four multi-draft argumentative writing tasks as take-home assignments. However, there was no explicit instruction on vocabulary or reading, and the course primarily focused on academic writing.

A total of 62 students from three intact classes were initially recruited: one DMC group (*N* = 20) and two classes for the traditional writing condition (*N* = 22 and *N* = 20, respectively; total *N* = 42), assigned to either the DMC or the traditional writing condition as whole-class units. Each group completed a background questionnaire in the first week. The research procedure, including all instructional interventions and assessments, was conducted over a period of five weeks.

The mean age of the included students was 19.71 (*SD* = 1.55), and there were 33 females and three males. The gender distribution reflected the institution's typical population of English majors. All students had studied English through the standard Chinese K-12 school curriculum, with only one student having lived for one year in an English-speaking country. Students in the DMC group all reported experience with digital tools commonly used for video essays, such as Microsoft PowerPoint, Canva, and iMovie.

During data analysis, data from some students were excluded. First, students who did not submit the final writing task or complete the posttest (DMC *N* = 3, traditional writing *N* = 19) were excluded. Additionally, in the traditional writing group, students whose final drafts contained more than 30% AI-generated content, as identified by both the instructors and an online AI detection tool, were excluded (*N* = 4). As a result, a total of 36 students were included in the analyses (DMC *N* = 17; traditional writing *N* = 19).

3.2. Research design and data collection procedure

All three classes (one DMC and two traditional writing classes) followed the same objectives and instructional plan, the only difference being the format of one argumentative writing task: a video essay for the DMC group and a monomodal text-based essay for the traditional writing group. One instructor taught the DMC group, and the other instructor taught the traditional monomodal writing group. To maintain curricular consistency, the two instructors followed the same course objectives and met regularly to discuss the course content and ensure alignment across groups.

Data were collected during the spring 2025 semester over five weeks. On the first day of the semester, students completed a writing proficiency test. In Week 2, they were introduced to the main writing assignment, with instructions relevant to their assigned modality. In the same week, they completed a vocabulary pretest to assess baseline knowledge of key lexical items in the upcoming reading

material and were then given the source article. In Week 3, students in the DMC group viewed a sample argumentative video essay to become familiar with the expectations and multimodal features of the task, whereas the traditional writing group read a sample argumentative essay and learned how to create a written outline prior to drafting. Both groups then began working on their assignments, with the DMC group outlining and planning storyboards and the traditional group outlining and drafting. During Week 4, all students submitted drafts of their work for feedback before completing their final versions. The first author, who has experience in DMC research and L2 writing instruction, provided feedback to the DMC group. The amount of feedback was controlled, with each student receiving between five and seven comments. Because of the limited feedback provided, language-related aspects were not the main focus. Instead, the feedback emphasized content, organization, and multimodal use. For the traditional writing groups, the first and second authors provided feedback to each of the two class groups. As with the DMC group, the amount and type of feedback were controlled. Based on the feedback they received, the DMC group continued developing their video projects using digital tools, while the traditional group revised their written essays. Both groups completed their final products as take-home assignments. In Week 5, all students submitted their final projects and immediately completed the post-task vocabulary quiz and a reading comprehension test to assess their vocabulary learning and reading comprehension. The timeline of the procedure is summarized in Figure 1.

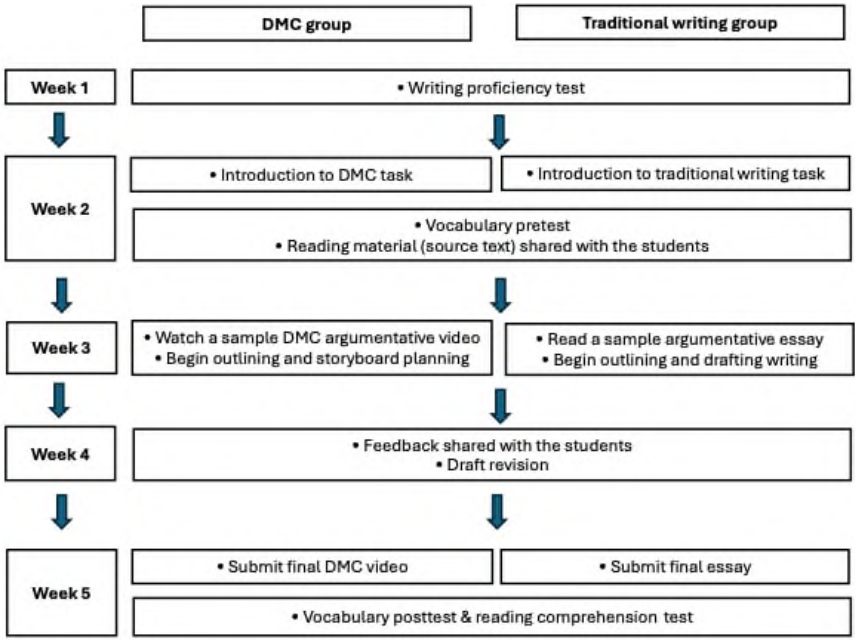

Figure 1 Data collection procedure

3.3. Materials

This section details the materials used in the current study, including a writing proficiency test, the reading material, the main writing task, vocabulary pretest and posttest, and a reading comprehension test. The grading rubrics for both DMC and traditional writing tasks are also described.

3.3.1. L2 writing proficiency test

To assess comparability of writing proficiency between the two groups, all students took an L2 writing proficiency test in the first week of the semester. The test was a timed writing task, with students given 45 minutes to complete an argumentative essay of at least 200 words in which they were asked to summarize a source text titled "With intelligent machines to do the thinking, will our brains get lazy?" and comment on whether they agreed or disagreed with the main claim.

A holistic rubric for the TOEFL integrated writing task was used to evaluate their baseline L2 writing proficiency, with a score ranging from 0 to 5. Components irrelevant to the current study context (e.g., the use of listening input) were excluded from scoring. The first and second authors scored the writing proficiency test. To assess interrater reliability, 10 essays were randomly selected and scored by the two raters in the first stage of scoring. The resulting scores indicated a high level of scoring consistency (r = .82). Any disagreements were resolved through discussion. The remaining essays were scored by the first author. An independent-samples t test was then conducted to compare the writing scores between the two groups. The results indicated no statistically significant difference between the groups (t = 0.566, p = .576). These scores were used as a covariate when answering the research questions.

3.3.2. Reading material and the main writing task

As part of the regular curriculum, students read an English article on a critical issue in contemporary society. The topic of generative artificial intelligence (AI) was selected based on the instructor's recommendation because of its timeliness and societal importance. Students were then given the article from *National Geographic* titled "Your biggest AI questions, answered" (Dhanesha & Locke, 2024) as a source text. The article addresses four problems associated with AI: the reliability of AI-generated or edited photos and images, hallucination, environmental impact, and the reinforcement of biases. The article contained 1,501 words

and had a Flesch-Kincaid grade level of 15.46. This commonly used readability index is based on sentence length and syllable count, and the score indicated that the text was appropriate for university-level EFL students.

Participants were subsequently required to choose one of the four AI-related issues presented in the source article, describe the issue, state whether they agreed or disagreed with the author's perspective, and propose one action to address the potential challenges associated with the issue. Students in the DMC group were asked to create a short video incorporating multiple modes of communication, while those in the traditional writing group were instructed to write a text-based essay. The specific instructions for each group are presented below and summarized in Table 1.

> You are going to apply for a job in a company which is specialized in various AI-related issues. One of their job application requirements is for you to share your opinion on the use of AI. You will review an article that describes four AI-related issues. In your video/essay, you will: (1) choose one of the four AI issues from the article and describe it, (2) explain whether you agree or disagree with the author's point, and (3) suggest one action for the problem to address its potential challenges.

Table 1 Writing directions

DMC	Traditional writing
Produce a 2- to 4-minute video based on your storyboard planning sheet using video-editing tools. In the video, various modes (e.g., pictures, video clips, soundtracks, color effects) are needed to complement your evidence or support your claim. You should pay attention to multimodal design, content, organization, and task fulfillment.	Compose an essay that is at least 300 words long, with an introduction, body paragraphs, and a conclusion.

3.3.3. Rubric for DMC and traditional writing

The writing outcomes were assessed based on an analytic rubric adapted from Cho and Kim (2024), focusing on content, organization, and language use (see Appendix A for the full rubric). To ensure task alignment, we modified the descriptors of the original rubric, which was designed for storybook summary-reflection writing, to fit our specific context: reading a source article and completing an argumentative task. The rubric assessed three main categories: content (source text integration, quality of the argument, and development of ideas), organization (structural coherence), and language use (accuracy and comprehensibility), with five points per category (maximum score of 15).

To facilitate comparison of the two task modes, the rubric used the same categories for both. For the DMC group, the use of multiple modes was included as an element in each category. Following Cho and Kim (2024), this study used the same analytic rubric to assess DMC and traditional writing in a reading-to-write task. The current study also considered the use of individual modes as well as "the orchestration of different modes" (p. 10) as crucial aspects of DMC outcomes. Thus, the use of different modes and the ways they were incorporated into the task were evaluated as part of each criterion.

The first and second authors assessed the final versions of all DMC videos and traditional essays. During the initial scoring stage, all DMC samples as well as two traditional writing samples were independently scored by the first and second authors. The scores for each category were compared to check for consistency. Interrater reliability across all samples was moderately high for all three components: content (r = .84), organization (r = .85), and language use (r = .83). Any disagreements were then resolved through discussion. Following calibration, the two raters assessed all remaining samples independently. The average scores were then used for further analyses.

3.3.4. Vocabulary pretest and posttest

In Week 2 of the semester, students completed a vocabulary pretest to check their familiarity with key terms in the source text and to determine the key vocabulary items to be assessed later in the posttest. The vocabulary pretest measured students' recognition of 30 content words selected from the source text by the first author based on their relevance to the article's themes and the likelihood that they would be unfamiliar to the learners. As the purpose of the pretest was to determine which words were unfamiliar to participants, the format was as follows:

> Do you know the meaning of this English word? "*hallucination*" (Yes/No)
> If yes, write the meaning in Chinese: _____ If yes, also write an English synonym: _____

Based on the test results, 15 words were identified as having familiarity rates below the mean of 58%. These 15 less-familiar words were selected for inclusion in the posttest vocabulary quiz (see Appendix B for the list of target vocabulary words).

The vocabulary posttest was administered soon after students submitted their final version of the task. To minimize the lag between task submission and the posttest, the deadline for the writing task was set 10 minutes before the class began, and the posttest was conducted immediately at the beginning of the class session. The vocabulary posttest included a total of 20 words from the

source text: 15 target words and five filler words. The vocabulary posttest followed the Vocabulary Knowledge Scale (VKS) developed by Wesche and Paribakht (1996), which allowed students to self-report their familiarity with and knowledge of each vocabulary item. VKS categories ranged from one to five as follows:

I) I don't remember having seen this word before
II) I have seen this word before but I don't know what it means
III) I have seen this word before and I think it means _____ (synonym or translation)
IV) I know this word. It means _____ (synonym or translation)
V) I can use this word in a sentence. For example: _____

The scoring for vocabulary knowledge was based on the above categories, taking both self-reported word knowledge and demonstrated vocabulary knowledge into consideration (see Table 2 for scoring guidelines). While Categories I and II focused on word recognition, Categories III through V required semantic and grammatical accuracy. In the scoring process, any incorrect semantic responses in the higher categories were lowered to a score of 2. For Categories III and IV, a score of 3 was assigned if a correct synonym or translation was provided. In Category V, if the meaning of the word was correctly provided but the sentence was incorrect, scores were adjusted between 3 and 4 based on the level of accuracy. The internal consistency of the 15 VKS items was assessed using Cronbach's alpha, which yielded a coefficient of .892, indicating high reliability of the measure.

Table 2 Vocabulary knowledge scale (VKS) scoring categories

Self-report category	Student response	Final score	Scoring criterion
I	Not familiar at all	1	No recognition of the word
II	Recognizes the word, but the meaning is not recalled	2	Word recognized, but meaning not provided
III	L1 (Chinese) meaning or L2 (English) synonym was provided	2	L1 meaning or L2 synonym was provided but incorrect
		3	Correct L1 meaning or L2 synonym provided
IV	L1 meaning or L2 synonym was provided	2	L1 meaning or L2 synonym was provided but incorrect
		3	Correct L1 meaning or L2 synonym provided
V	The word is used in a sentence with both semantic appropriateness and grammatical accuracy	2	L1 meaning or L2 synonym was provided but incorrect
		3	Correct L1 meaning or L2 synonym provided, but the sentence is grammatically and semantically incorrect
		4	Correct meaning and semantically appropriate sentence, but with grammatical errors
		5	Uses the word correctly in both meaning and grammar

3.3.5. Reading comprehension test

On the same day as the vocabulary posttest, students also took a reading comprehension test, which consisted of 10 multiple-choice questions and a summary-writing

task designed to assess students' understanding of the main ideas in the source text (see Appendix C). Multiple-choice questions were used to assess recognition, and summary writing was employed to measure higher-order synthesis and recall. The multiple-choice questions were balanced across the four main issues discussed in the original article, with two questions for each of the four AI-related problems and two additional questions addressing the overall synthesis of the article. To examine internal consistency, Cronbach's alpha was calculated. Cronbach's alpha coefficient for the ten multiple-choice items was .641.

For summary writing, students were asked to write a summary of the source text in five to seven sentences, approximately 100 words. The summaries were written in Chinese to minimize the influence of students' L2 proficiency. Two raters, the second author and a graduate student in applied linguistics, both native Chinese speakers, scored the summaries. The summaries were scored based on idea units, with one point given for each of the four main problems discussed in the article and one point for overall completeness (total of 5 points). Interrater reliability was high (r = .964), and average scores were used for the main analyses to answer RQ3.

3.4. Data analysis

All statistical analyses were performed using R version 4.3.3 (R Core Team, 2024). For all three research questions, analyses of covariance (ANCOVAs) were conducted to control for baseline writing proficiency and examine the effect of composition modality on L2 learning. In all ANCOVA models, the independent variable was composition modality (DMC vs. traditional writing), and writing proficiency test scores were included as a covariate to account for differences in baseline writing ability. For RQ1, three ANCOVA models were conducted, with content, organization, and language-use subscores for the main writing task serving as the respective dependent variables. Regarding RQs 2 and 3, dependent variables were posttest scores from the VKS (RQ2) and reading comprehension measures (multiple-choice and summary writing; RQ3). Partial eta squared (η^2_p) values were reported as measures of effect size.

4. Results

This section reports comparisons of writing scores, the vocabulary posttest scores, and reading comprehension posttest scores.

4.1. Writing score comparison between DMC and traditional writing

Table 3 presents descriptive statistics for participants' scores on the writing proficiency test. The two groups obtained similar scores (*M* = 3.65 for the DMC group and *M* = 3.47 for the traditional writing group). Based on the TOEFL holistic rubric criteria, these scores indicated that students in both DMC and traditional writing groups were at an intermediate level of L2 writing proficiency. Table 4 presents descriptive statistics for the three trait scores on the main writing task across the two groups, and Figure 2 shows the corresponding boxplots.

Table 3 Descriptive statistics for writing proficiency test scores

	Composition modality	*N*	*M*	*SD*	*Min*	*Max*
Writing proficiency test	DMC	17	3.65	0.86	2	5
	Traditional	19	3.47	0.96	2	5

Note. The total possible score was 5

Table 4 Descriptive statistics of main task writing scores

	Composition modality	*N*	*M*	*SD*	*Min*	*Max*
Content	DMC	17	3.68	0.86	2.5	5
	Traditional	19	3.32	0.63	2	4.5
Organization	DMC	17	3.91	0.69	3	5
	Traditional	19	3.39	0.66	2	5
Language use	DMC	17	3.65	0.68	3	5
	Traditional	19	3.61	0.79	1	4.5
Total	DMC	17	11.24	1.99	9	15
	Traditional	19	10.32	1.58	6.5	13

Note. The total possible score for both types of writing was 15

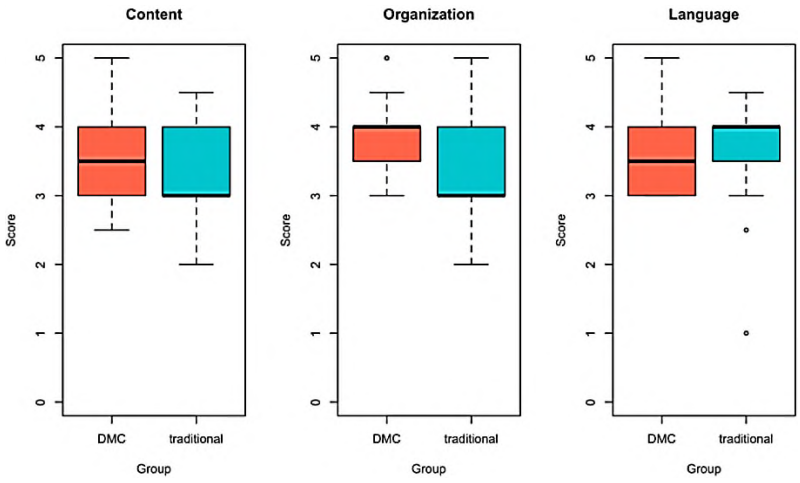

Figure 2 Boxplots of writing scores across the two groups (DMC vs. traditional)

Overall, students in the DMC group scored higher on all subcategories, resulting in a difference of approximately 1 point in the total writing score. The ANCOVA revealed a significant difference between the two groups in organization (F (1, 33) = 4.94, p = .033, η^2_p = 0.13). However, there were no significant differences between the two groups in terms of content (F (1, 33) = 1.87, p = .18, η^2_p = 0.05) or language use (F (1, 33) = 0.01, p = .924, η^2_p = .001). These findings suggest that the DMC group outperformed the traditional group primarily in terms of organizational quality, while performance in content and language use did not differ significantly.

4.2. Effects of composition modality on vocabulary learning

Table 5 presents descriptive statistics for the vocabulary posttest scores measured using the VKS. For the total VKS scores, the DMC group had a slightly higher mean score (M = 35.88, SD = 11.56) than the traditional writing group (M = 32.63, SD = 8.69).

Table 5 Descriptive statistics of total VKS scores

Composition modality	*M*	*SD*	*Min*	*Max*
DMC (*N* = 17)	35.88	11.56	22	66
Traditional (*N* = 19)	32.63	8.69	18	53

Note. The total possible score was 5 (VKS level) × 15 (number of vocabulary items) = 75. VKS = Vocabulary Knowledge Scale

The ANCOVA, which controlled for writing proficiency, showed that the group difference in VKS scores was not statistically significant (F(1,33) = 1.51, p = .228, η^2_p = .04). To further understand the score distributions, individual word-level results were analyzed (see Table 6). The total number of word entries in Table 6 was calculated by multiplying the number of participants in each group (N = 17, N = 19) by the 15 target words. The percentages represent the proportion of word entries in each VKS category within each group.

Table 6 Word entries for each VKS score per group

			Number of words for VKS scores (%)				
Composition modality	*N*	Total number of word entries	1	2	3	4	5
DMC	17	255	57 (22.35%)	129 (50.59%)	16 (6.27%)	18 (7.06%)	35 (13.73%)
Traditional	19	285	75 (26.32%)	150 (52.63%)	14 (4.91%)	27 (9.47%)	19 (6.67%)

Note. The meaning of each VKS score is listed in Table 2

Each student was assessed on 15 target words, resulting in 255 word entries for the DMC group (17 students × 15 words) and 285 for the traditional group (19 students × 15 words). Across both groups, the proportion of responses receiving a score of 1 (i.e., no knowledge of the word) was relatively similar, accounting for about a quarter of all word entries in each group. A score of 2 (i.e., recognition without recall of meaning) was the most common, observed for about half of all entries in both groups. Scores of 3, which reflect successful meaning recall but limited productive use, were less frequent overall, representing approximately 5% of all word entries in both groups. Scores of 4 and 5 reflect productive knowledge, with a score of 4 indicating semantically accurate use of the word in a sentence (despite possible grammatical errors), and a score of 5 reflecting both semantically and grammatically correct usage. As shown in Table 6, although both groups had similar proportions of lower scores (1 and 2), which do not demonstrate learners' learning of new words, the DMC group showed a slightly higher percentage of responses receiving a score of 5 (13.73% vs. 6.67%), suggesting students in the DMC group were somewhat more capable of using the target words accurately and grammatically in context.

4.3. Effects of composition modality on reading comprehension

Table 7 presents descriptive statistics for participants' scores on the reading comprehension task. The DMC group had a slightly higher mean score on both reading comprehension measures (M = 7.47, SD = 1.38 for multiple-choice; M = 3.32, SD = .88 for summary) than the traditional writing group (M = 6.47, SD = 2.39 for multiple-choice; M = 2.55, SD = 1.19 for summary). Figure 3 shows boxplots of reading comprehension scores by composition modality.

Table 7 Descriptive statistics of reading comprehension posttest

	Composition type	*N*	*M*	*SD*
Multiple-choice questions	DMC	17	7.47	1.38
	Traditional	19	6.47	2.39
Summary writing	DMC	17	3.32	0.88
	Traditional	19	2.55	1.19

Note. The total possible score was 10 for multiple-choice questions and 5 for summary writing

To examine whether students' reading comprehension differed between the groups, two ANCOVAs were conducted, controlling for writing proficiency scores. For the multiple-choice reading comprehension measure, there was no significant effect of group ($F(1, 33) = .03$, $p = .860$, $\eta^2_p = .001$). However, for the summary-writing measure, group had a significant effect, ($F(1, 33) = 4.32$, $p = .045$, $\eta^2_p = .12$), with the

DMC group outperforming the traditional group. Overall, the DMC group had higher mean scores on both reading comprehension measures, although the between-group difference was significant only for the summary-writing task.

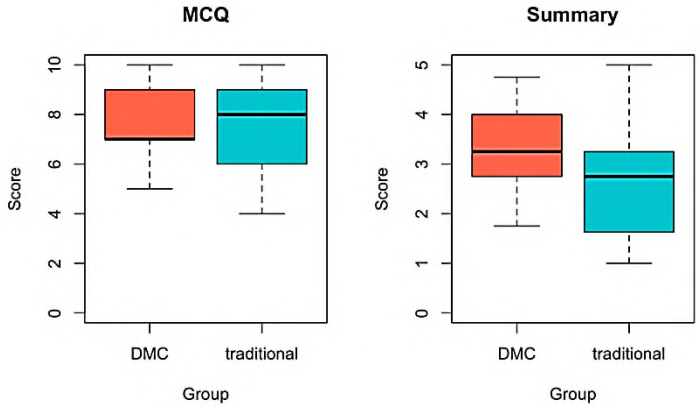

Figure 3 Boxplots of reading comprehension scores by composition modality

5. Discussion

This quasi-experimental study compared the effects of DMC and traditional monomodal writing in an integrated reading-to-write task. Regarding RQ1, the results showed that the DMC group outperformed the traditional writing group in organization, but no significant differences were found in content and language use. These results partially align with previous findings (e.g., Cho, 2022; Cho & Kim, 2024; Kim & Belcher, 2020), which showed no significant differences in writing quality between DMC and traditional writing groups. However, the majority of these studies employed within-participant comparisons, examining how the same learners performed across different composing modes, and learners' initial writing proficiency was not taken into consideration. In contrast, the current study compared two groups of learners while accounting for writing proficiency test scores. The finding from RQ1 indicates that, when their writing proficiency was held constant, learners performed similarly across the two modalities in content and language use, consistent with previous within-participant findings on language use.

The finding that the DMC group performed better in organization suggests that DMC can help learners structure their ideas more effectively, possibly because of the storyboard-planning stage and the meaning-making processes involved in coordinating multiple semiotic resources. This aligns with the notion that DMC fosters multimodal communicative competence through learners' active orchestration of various modes, "arranging available meaning-making resources into a

multimodal whole, making authorial decisions appropriately for specific audiences and purposes" (Shin et al., 2020, p. 2). In addition, our finding that the DMC group performed better in organization may provide further support for the benefits of storyboarding as a guided planning tool because it helps L2 writers to explicitly map, sequence, and visualize the relationships among different components scene by scene, a process that may require more rigorous structural planning than a traditional text-only outline (Kim et al., 2022). Furthermore, the sample video provided prior to the task likely offered students a structural template for how to organize these elements.

With respect to RQ2, which examined the effects of composition modality on vocabulary acquisition after controlling for baseline writing proficiency, the ANCOVA model showed that vocabulary posttest scores did not differ significantly between the two groups. This finding is in line with Yang and Wu (2012), who also reported no significant differences in vocabulary learning outcomes between digital and text-based writing conditions. However, compared to the traditional writing group, the DMC group showed a higher proportion of words used productively and accurately at the sentence level (VKS score 5), suggesting a possible advantage of DMC in promoting semantically and grammatically accurate use of the vocabulary items from the source text. This descriptive pattern is consistent with the levels of processing framework (Craik & Lockhart, 1972), suggesting that the active resemiotization of word meanings into visual and auditory modes may require deeper processing, thereby leading to better retention. According to Mayer's (2001) cognitive theory of multimedia learning and Paivio's (1986) dual coding theory, this resemiotization process allows learners to build stronger mental representations by linking linguistic items with nonverbal imagery, facilitating better retention compared to text-only writing. One possible interpretation is that, when engaging in DMC, students actively evaluate and select visual or auditory representations to match those words. This additional cognitive involvement may have promoted deeper semantic processing, supporting vocabulary retention without distracting from linguistic processing.

Regarding reading comprehension (RQ3), results of the ANCOVAs showed that group differences were found in summary writing but not in multiple-choice comprehension. The nonsignificant multiple-choice result is in line with Marino (2025) but contradicts Yang and Wu (2012), who found that the DMC group performed significantly better in multiple-choice reading comprehension than the traditional writing group. This discrepancy may be explained by the collaborative nature of the DMC task in Yang and Wu (2012), as peer interaction may have supported comprehension through discussion and shared meaning-making. In contrast, the DMC task in the present study was completed individually, meaning that comprehension relied solely on learners' independent engagement with the text.

Moreover, the nonsignificant difference in multiple-choice scores in our study may indicate that both composition modalities can support general reading comprehension when the focus is on recognition and recall.

However, when learners were required to summarize the text by integrating main ideas and expressing them in their own words, the DMC group performed better. This pattern suggests that DMC may foster deeper conceptual understanding and idea integration rather than simple recall, a finding that may be linked to learners' extended engagement during DMC processes, as shown in previous research. As Zalazar (2025) highlights, the process of DMC requires more time on task than text-only writing. Such engagement may encourage learners to reflect more carefully on the content of their output as they transform textual information into visual, auditory, and spatial representations. From the perspective of multiliteracies pedagogy, our findings suggest that DMC may enable learners to engage in deeper meaning-making processes while reading a source text, encouraging them to actively reproduce meaning. This process of transforming textual input from the source text into a multimodal ensemble, known as resemiotization, may require greater synthesis and rhetorical decision-making than traditional writing. Consequently, this resemiotization process may lead to deeper task engagement and stronger performance in summary writing, which demands higher-order skills such as synthesizing and integrating ideas rather than simply recalling them.

6. Conclusion

This study is one of the few studies that investigate learning outcomes associated with reading-to-write DMC tasks, and it focused specifically on the impact of composition modality on writing quality, vocabulary learning, and reading comprehension. Overall, the results of the current study indicate that DMC and traditional writing conditions can both support L2 learning in integrated writing instruction. When learners' writing proficiency was held constant, the L2 performance patterns of the DMC group were similar to those of the traditional writing group. Meanwhile, DMC may offer additional benefits for productive vocabulary use and source-text synthesis. It is also important to note that the instructional focus was primarily on writing, as the course objective was to develop students' academic writing skills. Reading materials served only as support for writing. Nonetheless, DMC showed potential benefits for productive vocabulary use and summary writing, even without explicit reading and vocabulary instruction.

From a theoretical perspective, the study extends the multiliteracies framework to integrated reading-to-write tasks, showing how multimodal resources can support the development of multiple L2 literacy skills. In light of multiliteracies

pedagogy (New London Group, 1996), our findings offer empirical support for DMC in L2 classrooms. These findings suggest that DMC and traditional monomodal writing can both benefit L2 learning in integrated tasks. Therefore, DMC can be incorporated into reading-to-write instruction as a way to support textual engagement, organization skills, and integrated literacy development. Additionally, engaging in DMC may support higher-order cognitive skills necessary for academic success, such as structural organization and global text synthesis.

The study is not without limitations, which should be taken into consideration when interpreting the findings and designing future research. First, because of a relatively small sample size and the use of a single task, the findings may not be generalizable. Second, there are several potential confounding variables. For example, students completed the writing task outside of class, which limits insights into how students engaged in the task or how much time they actually spent on it. Controlled, laboratory-based studies could provide more precise insights. Furthermore, the different pre-task planning conditions (storyboarding for the DMC group versus text-based outlining for the traditional group) may have provided unequal scaffolding. As storyboarding may provide rigorous structural planning opportunities, this difference may have acted as a confounding factor, particularly influencing the organizational scores. The final limitation is related to the absence of process-oriented insights. Future studies should investigate both writing processes and products, taking a learner-centered approach. For instance, data such as classroom observations or stimulated-recall interviews could be collected along with the outcome data. We also hope that future researchers extend this line of inquiry by exploring how L2 learners engage with multimodal composing across varied task types and genres (e.g., narrative, summary, and expository).

References

Ajayi, L. (2008). Meaning-making, multimodal representation, and transformative pedagogy: An exploration of meaning construction instructional practices in an ESL high school classroom. *Journal of Language, Identity, and Education, 7(*3-4), 206-229. https://doi.org/10.1080/15348450802237822

Akoto, M., & Li, M. (2025). Exploring collaborative digital multimodal composition: Interaction patterns and texts in the French as a foreign language context. *System, 129,* 103583. https://doi.org/10.1016/j.system.2024.103583

Casanave, C. P. (2017). *Controversies in second language writing: Dilemmas and decisions in research and instruction* (2nd ed.). University of Michigan Press.

Cimasko, T., & Shin, D. S. (2017). Multimodal resemiotization and authorial agency in an L2 writing classroom. *Written Communication, 34*(4), 387-413. https://doi.org/10.1177/0741088317727246

Cho, H. (2022). *Extensive reading-to-write task outcome, vocabulary learning and Korean EFL students' perceptions towards task performance: Comparing monomodal and multimodal composing tasks* [Doctoral dissertation, Georgia State University].

Cho, H., & Kim, Y. (2024). Comparing the characteristics of EFL students' multimodal composing and traditional monomodal writing: The case of a reading-to-write task. *Language Teaching Research*, 28(6), 2487-2512. https://doi.org/10.1177/13621688211046740

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cumming, A. (2014). Assessing integrated skills. In J. Kunnan (Ed.), *The companion to language assessment: Volume I: Ability, contexts, and learners* (pp. 216-229). Wiley.

Dhanesha, N., & Locke, C. (2024, October 15). Your biggest AI questions, answered. *National Geographic*. https://www.nationalgeographic.com/science/article/ai-effects-questions-concerns

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *Canadian Modern Language Review, 58*(4), 599-620. https://doi.org/10.3138/cmlr.58.4.599

Hafner, C., & Miller, L. (2011). Fostering learner autonomy in English for science: A collaborative digital video project in a technological learning environment. *Language Learning & Technology, 15*(3), 68-86. https://doi.org/10.64152/10125/44263

Hirvela, A. (2016). *Connecting reading and writing in second language writing instruction* (2nd ed.). University of Michigan Press.

Jiang, L. (2017). The affordances of digital multimodal composing for EFL learning. *ELT Journal, 71*(4), 413-422. https://doi.org/10.1093/elt/ccw098

Jiang, L., & Hafner, C. (2025). Digital multimodal composing in L2 classrooms: A research agenda. *Language Teaching, 58*(4), 528-546. https://doi.org/10.1017/S0261444824000107

Jiang, L., & Luk, J. (2016). Multimodal composing as a learning activity in English classrooms: Inquiring into the sources of its motivational capacity. *System, 59*, 1-11. https://doi.org/10.1016/j.system.2016.04.001

Jung, J. (2020). Effects of content support on integrated reading-writing task performance and incidental vocabulary learning. *System, 93*, 102288. https://doi.org/10.1016/j.system.2020.102288

Kessler, M. (2024). *Digital multimodal composing: Connecting theory, research and practice in second language acquisition*. Multilingual Matters.

Kim, Y., & Belcher, D. (2020). Multimodal composing and traditional essays: Linguistic performance and learner perceptions. *RELC Journal, 51*(1), 86-100. https://doi.org/10.1177/0033688220906943

Kim, Y., Belcher, D., & Peyton, C. (2023). Comparing monomodal traditional writing and digital multimodal composing in EAP classrooms: Linguistic performance and writing development. *Journal of English for Academic Purposes, 64,* 101247. https://doi.org/10.1016/j.jeap.2023.101247

Kim, Y., Kang, S., Nam, Y., & Skalicky, S. (2022). Peer interaction, writing proficiency, and the quality of collaborative digital multimodal composing task: Comparing guided and unguided planning. *System, 106*, 102722. https://doi.org/10.1016/j.system.2022.102722

Knoch, U., & Sitajalabhorn, W. (2013). A closer look at integrated writing tasks: Towards a more focused definition for assessment purposes. *Assessing Writing, 18*(4), 300-308. https://doi.org/10.1016/j.asw.2013.09.003

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

Lee, S. Y., Lo, Y. H. G., & Chin, T. C. (2021). 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

Li, M., & Pham, Q. N. (2025). Three heads are better than one? Digital multimodal composition completed collaboratively versus individually. *Language Teaching Research, 29*(5), 1866-1888. https://doi.org/10.1177/13621688221102536

Mai, S., Ma, S., & Pei, H. (2023, July). A comparative study on the effectiveness of multimodal input and digital multimodal composing on vocabulary learning. In *2023 International Symposium on Educational Technology (ISET)* (pp. 244-248). IEEE. https://doi.org/10.1109/ISET58841.2023.00055

Marino, F. (2025). *Digital multimodal composing in an L2 Italian classroom: The effects of digital video projects on students' development and investment* [Doctoral dissertation, University of South Florida].

Mayer, R. E. (2001). *Multimedia learning.* Cambridge University Press.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, 66, 60-92. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Nishioka, H. (2016). Analysing language development in a collaborative digital storytelling project: Sociocultural perspectives. *System, 62*, 39-52. https://doi.org/10.1016/j.system.2016.07.001

Paivio, A. (1986). *Mental representations: A dual-coding approach*. Oxford University Press.

Payant, C., McDonough, K., Uludag, P., & Lindberg, R. (2019). Predicting integrated writing task performance: Source comprehension, prewriting planning, and individual differences. *Journal of English for Academic Purposes, 40*, 87-97. https://doi.org/10.1016/j.jeap.2019.06.001

Qu, W. (2017). For L2 writers, it is always the problem of the language. *Journal of Second Language Writing, 38*, 92-93. https://doi.org/10.1016/j.jslw.2017.10.007

R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.3.3) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

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, 47,* 100714. https://doi.org/10.1016/j.jslw.2020.100714

Shin, S. Y., & Ewert, D. (2015). What accounts for integrated reading-to-write task scores? *Language Testing, 32*(2), 259-281. https://doi.org/10.1177/0265532214560257

Vandommele, G., Van den Branden, K., Van Gorp, K., & De Maeyer, S. (2017). In-school and out-of-school multimodal writing as an L2 writing resource for beginner learners of Dutch. *Journal of Second Language Writing, 36*, 23-36. https://doi.org/10.1016/j.jslw.2017.05.010

Wesche, M., & Paribakht, T. S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53*(1), 13-40. https://doi.org/10.3138/cmlr.53.1.13

Xu, Y. (2023). Investigating the effects of digital multimodal composing on Chinese EFL learners' writing performance: A quasi-experimental study. *Computer Assisted Language Learning, 36*(4), 785-805. https://doi.org/10.1080/09588221.2021.1945635

Yang, Y. T. C., & Wu, W. C. I. (2012). Digital storytelling for enhancing student academic achievement, critical thinking, and learning motivation: A year-long experimental study. *Computers & Education, 59*(2), 339-352. https://doi.org/10.1016/j.compedu.2011.12.012

Zalazar, M. L. (2025). Digital multimodal composing versus traditional monomodal writing: A look at the Spanish imperfect subjunctive from a process and product perspective. *System*, *132*, 103718. https://doi.org/10.1016/j.system.2025.103718

APPENDIX A

Analytic rubric for the writing task

Score	Content	Organization	Language use
5 Excellent	• Effectively addresses the topic and task with a clearly presented central argument, integrating relevant information from the article; demonstrates effective integration of textual, audio, visual effects, and author narration to represent the topic (where applicable) • Provides an opinion statement and suggests an action to address the challenge(s) mentioned in the article with appropriate details; shows effective use of multimodal resources to strengthen explanation and support (where applicable)	• Very clearly organized and fully developed with effective and relevant topic, support, and conclusion • Always maintains a consistent point of view and focus; clear and effective cohesion across ideas and modes (text + visuals + audio) (where applicable)	• Consistently demonstrates varied sentence structures and contextually appropriate vocabulary, with only minor errors that do not impede comprehensibility • Oral narration (where applicable) is clear, fluent, and easily comprehensible
4 Good	• Addresses the topic and task well with generally relevant and accurate information from the article, though some points may lack full clarity or detail; generally demonstrates effective integration of textual, audio, visual effects, and author narration to represent the topic • Provides an opinion statement and suggests a solution, supported with mostly appropriate details; generally shows effective use of multimodal resources to support explanation	• Mostly well organized and well developed and clarifies most relationships among ideas, though there may be some occasional redundancy or unclear connections • Mostly maintains a consistent point of view and focus; shows mostly effective integration of modes	• Mostly demonstrates a variety of sentence structures and contextually appropriate vocabulary, though there are occasional minor errors that do not severely interfere with meaning • Oral narration is mostly clear and comprehensible, with minor hesitations or errors
3 Developing	• Addresses the topic and task with some information from the article, but with some inaccuracies, vagueness, or generalizations; occasionally uses textual, audio, visual effects, and author narration to represent the topic • Provides somewhat developed opinion statement and solution with elaborations or ideas that may be superficial; shows occasional uses of multimodal resources to support explanation	• Somewhat organized and developed, though connections among ideas may be occasionally unclear • Occasionally maintains a consistent point of view; demonstrates the integration of different modes, though they may include somewhat weak connections	• Occasionally uses contextually appropriate vocabulary, though sentence structure may be limited in variety and/or some errors may interfere with comprehensibility • Oral narration may be generally understandable, though it may be occasionally unclear or difficult to follow
2 Inadequate	• Limited development in response to the topic and task with unclear argument that includes only minimal or inaccurate information from the article; rarely uses textual, audio, visual effects, and author narration to represent the topic • Provides an inappropriate or insufficient opinion statement and/or solution; shows minimal and/or ineffective use of multimodal resources	• Demonstrates weak organization and coherence with poor connection across ideas • Rarely maintains a consistent point of view and focus; limited use of different modes with unclear connections between them	• Rarely uses varied sentence structures or range of words, with frequent errors that impede comprehensibility • Oral narration is often unclear or difficult to understand
1 Poor	• Seriously underdeveloped content without any relevant information from the article, or information is seriously inaccurate • Provides little or no opinion statement and/or solution; no use of multimodal resources	• Contains serious disorganization and lack of coherence • Little or no connections between ideas; no attempt to integrate different modes	• Uses only limited sentence structures and contextually inappropriate vocabulary, with serious and frequent errors resulting in significant difficulties in comprehensibility • Oral narration is incomprehensible or absent

APPENDIX B

List of target vocabulary words

Word (part of speech)
1. radically (adv)
2. underlying (adj)
3. notorious (adj)
4. bias (n)
5. manipulation (n)
6. generative (adj)
7. marginalize (v)
8. hallucinate (v)
9. confront (v)
10. flawed (adj)
11. fallible (adj)
12. cryptographic (adj)
13. weigh in (v)
14. plausible (adj)
15. replicate (v)
16. command (v)
17. compound (n)
18. prioritize (v)
19. made up (v)
20. desired (adj)

APPENDIX C

Reading comprehension posttest

Section 1. Multiple-choice questions

Circle the correct answer for the questions below. There is only one answer for each question.

1. What is Fred Ritchin's view on AI-generated images?
   A) We should not trust images created by AI.
   B) AI-generated visuals make it harder to trust photographs.
   C) Setting clear rules for AI-generated images will completely solve the problem.
   D) Digital photography is more authentic than film photography.

2. What is true about the current requirements for National Geographic photographers?
   A) Policies on photo manipulation are being developed.
   B) Photographers are encouraged to use AI-generated images.
   C) Photographers are required to shoot photos in uncompressed and unprocessed format.
   D) Photographers don't have to report any photo manipulation they have made.

3. According to Rayid Ghani, why do AI models sometimes produce false but convincing statements?
   A) They focus on probability rather than truth.
   B) They are designed to make random guesses.
   C) They use outdated information only.
   D) They deliberately mislead users.

4. What does "hallucinate" mean in the context of AI?
   A) To intentionally produce incorrect responses
   B) To provide outdated information
   C) To misinterpret human input
   D) To generate false but plausible responses

5. What environmental concern is associated with AI?
   A) AI produces toxic chemicals.
   B) Renewable energy, such as solar power, cannot be used in AI systems.
   C) AI requires massive energy and water consumption.
   D) AI causes deforestation.

6. According to Shaolei Ren, why is energy use associated with AI a problem?
   A) The energy consumed for AI cannot be reused.
   B) Companies don't make efforts to control their energy systems.
   C) Collaboration between countries, such as the US and EU, faces challenges.
   D) Individuals cannot resist using AI in their daily and professional lives.

7. How does AI reinforce stereotypes, racism, and inequities?
   A) Recognizing indigenous knowledge and cultural values

B) Prioritizing perspectives and agendas of those in power
C) Making AI only use European-based data
D) Restricting AI development to large companies

8. What is the best way to make AI more representative of global perspectives?
A) Limiting AI's access to diverse datasets
B) Asking large tech companies to be more responsible
C) Encouraging local AI development in different communities
D) Reducing AI's influence in education

9. What is something that AI *can* do?
A) Check the accuracy of data
B) Identify bias and stereotype within given information
C) Searching its knowledge base for relevant details
D) Determine a reliable source from the internet

10. How do the experts suggest we can address the negative impacts of AI on society?
A) By shifting AI development to local communities and diversifying the voices involved in creating AI tools
B) By relying on AI to fix existing social issues
C) By allowing AI to make decisions on behalf of governments and companies
D) By banning AI in all forms of journalism

Section 2. Summary – Summarize the reading *Your biggest AI questions, answered* in about 100 words (5-7 sentences) in Chinese. Do not refer to the original reading at this point, but try to provide information you remember as much as you can.