

DEVELOPING SECONDARY SCHOOL STUDENTS' COMPETENCE IN UTILIZING GENERATIVE ARTIFICIAL INTELLIGENCE FOR MULTIMODAL TEXT WRITING

Nguyen The Hung

Faculty of Philology, Hanoi National University of Education, Hanoi city, Vietnam

*Corresponding author: Nguyen The Hung, e-mail: nthung@hnue.edu.vn

Received June 9, 2025. Revised July 28, 2025. Accepted December 9, 2025.

Abstract. This study aims to develop students' competence in utilizing generative artificial intelligence (GenAI) in writing multimodal texts in secondary schools. The study proposes a competency framework for GenAI application in students' multimodal text writing, consisting of three basic dimensions: Foundational Understanding, Effective Interaction with GenAI, and Strategic Application in the Multimodal Writing Process. Additionally, the study develops three specific pedagogical solutions for application in secondary school teaching activities. These measures aim to equip students with foundational knowledge about multimodal texts and GenAI, hone their skills in interacting with AI tools, and integrate GenAI into the stages of the multimodal text writing process from preparation, information gathering, brainstorming, outlining, and writing to post-writing editing. The ultimate goal is to help students master GenAI tools and use GenAI appropriately in the process of writing multimodal texts, ensuring that they meet the requirements for multimodal text writing skills outlined in the 2018 General Education Curriculum for Language Arts.

Keywords: Generative AI, multimodal text, competence, competency framework.

1. Introduction

Currently, many artificial intelligence tools such as ChatGPT, Gemini, and Midjourney are becoming increasingly popular and playing an important role in education. This presents teachers, students, and even education administrators with many prospects and obstacles. On the other hand, it is undeniable that GenAI is increasingly appearing in most areas of life in many diverse forms, not only language but also images, sounds, videos, and many other formats. Along with that, multimodal text and the ability to create multimodal text are also highly valued.

Parallel to technological development, the importance of multimodal text generation capabilities is gaining widespread acknowledgment. Multimodal text (in contrast to monomodal text) is text created by two or more modes (including language, images, sound, video, links, activities, interactions, etc.). These modes are closely integrated, demonstrating a certain degree of interaction, influence, and connection. They can be combined with digital support to form a unified whole, expressing complete meaning and fulfilling specific communication purposes. Information is therefore no longer conveyed solely through written language but is a complex combination of various channels of expression such as language, static and dynamic images, sound, spatial layout, and color [1].

In Vietnam, the 2018 General Education Curriculum for the subject of Literature sets specific requirements for developing students' multimodal writing skills at different grade levels, from grade 5 (writing reports using tables), grade 6 (summarizing texts using diagrams), or grade 9 (writing advertisements incorporating non-verbal media) [2]. These requirements highlight the significance of prioritizing the development of multi-modal writing skills. However, attention should also be paid to related issues such as AI ethics and how to use AI while maintaining the active role of both teachers and learners. These are urgent matters that need to be addressed to achieve the comprehensive development of learners [3].

Since GenAI entered our lives, many studies have focused on this topic as an effective tool in many fields, including education. Specifically in teaching, GenAI can effectively support learners in improving learning quality, assessment, personalizing learning paths, and helping teachers enhance their lesson planning skills and flexibly apply active teaching methods. In writing instruction, many studies have also shown that GenAI can support learners in various stages such as brainstorming, building a framework for writing, and assisting with writing and editing. Recent studies have confirmed the effectiveness of GenAI in improving students' writing quality [4]. However, researchers also warn that overuse of GenAI can lead to learners becoming lazy in their thinking, dependent on AI, and less capable of problem-solving. These concerns are well-founded because when overly guided by GenAI and using it habitually, learners are highly likely to become completely dependent on it. Additionally, some studies suggest that GenAI also impacts user privacy by sharing personal data on AI platforms [5].

The theory of multimodal texts and teaching multimodal texts has also attracted attention, notably from researchers such as Kress and van Leeuwen. In his research, Kress identified that the modes in multimodal texts must be closely related to each other in the overall context, creating a coherent system [6]. Modes in multimodal texts are divided into many different forms, basically divided into the following three groups: (1) Static modes, including words, still images (drawings, photographs, illustrations), graphics (diagrams, charts, schematics, maps), etc.; (2) Dynamic modes: sound (sound effects, music, sound effects, silence), animated images (animation), video (recorded images and sound from filming devices), links (links from one VB to another), real-world activities (through gestures, body language, spoken language, posture, and the acting of real-world characters), etc.; and (3) Interactive communication channel group: interactive programs that require feedback and interaction between the communication channel creator and the communication channel recipient or between communication channel recipients themselves (interactive videos, interactive images, seminars, forums, social networks, buying and selling websites, exchanges, etc.). It can be seen that research on the integration of these methods in teaching writing in secondary schools is still limited, having only stopped at the channel of language expression.

2. Content

2.1. Theoretical Basis

2.1.1. Generative Artificial Intelligence (GenAI)

GenAI is a concept that refers to the use of artificial intelligence to create new, creative, and unique products, including images, sounds, videos, and many other complex formats. Products created from artificial intelligence can be processed through a system of rules or programming, demonstrating the inevitable development of information technology. In relation to multimodal text, GenAI can also create different channels of expression (corresponding to the channels of expression in multimodal text), thus confirming that GenAI has great potential for application in creating multimodal text. Depending on the type of multimodal text that students need to create, the way GenAI is applied will change accordingly.

2.1.2. Multimodal Texts

Multimodal texts are defined as texts that combine two different expressive channels to create meaning. According to Kress and van Leeuwen, each expressive channel is multimodal in nature (for example, images include color, lines, and layout) [6]. Many other studies also argue that there is no such thing as a monomodal text, emphasizing that the boundaries between modes have become blurred over time. Whether the text is written, spoken, hyperlinked, or a mixture of these forms, it contains multiple modes within itself. Another view holds that multimodal texts must involve the combination of linguistic means and other means such as symbols, diagrams, charts, images, and sounds [7]. This view is more relevant to domestic education as it is agreed upon in the 2018 General Education Curriculum.

Multimodal modes in texts are complete communication tools such as language, images, sound, video, actions, links, interactions, etc. This is also the perspective used in the 2018 General Education Curriculum. Multimodal modes in texts have the following key characteristics: first, modes in multimodal texts are combined and integrated with each other. Language is only one mode; there are other modes such as images, actions, sounds, and many others. Readers need to know how they are created and understand the meaning of the integrated modes in multimodal texts [8]. Second, modes in multimodal texts demonstrate interaction and influence each other, but some modes play a more prominent role than others in expressing the meaning of multimodal texts. Third, modes in multimodal texts, especially in digital multimodal texts, are characterized by technical support, demonstrating flexibility in terms of time and place of reception. Unlike monomodal texts that depend on surface materials, non-printed multimodal texts and digital multimodal texts provide information to recipients easily, without being constrained by geographical boundaries or fixed time frames.

2.1.3. Requirements for Multimodal Text Writing in the 2018 GEC for Language Arts

The main goal of the 2018 General Education Curriculum for Language Arts is to develop students' qualities and competencies, focusing on three key skills: reading, writing, speaking, and listening. The program emphasizes flexibility, allowing educational institutions to proactively select appropriate teaching content, provided that it meets the required competency standards for each grade level, rather than teaching specific texts. When considering the requirement to create texts, specifically multimodal texts, the General Education Curriculum demonstrates its focus on developing the ability to create various types of multimodal texts, such as: Grade 5 – Requirement to write a work report using tables; Grade 6 – Requirement to summarize the content of certain texts using diagrams; Grade 9 – Requirement to write an explanatory essay about a scenic spot using diagrams, tables, and images, or requirement to write an advertising text using a combination of non-verbal means; Grade 10 – Requirement to write a research report using citations and footnotes; Grade 11 – Write a research report using citations, footnotes, and create a bibliography; Grade 12 – Write a project assignment report using a combination of different modes of expression such as citations, footnotes, a bibliography, and appropriate supporting media [2]. Thus, it can be seen that the content of creating multimodal texts is spread across all grade levels with increasing requirements. Students need to combine different modes to create corresponding multimodal texts. Therefore, teachers need to flexibly apply measures to develop students' multimodal text writing skills to meet the requirements of the 2018 General Education Curriculum.

2.2. Competency Framework for utilizing Generative AI in Multimodal Text Writing

2.2.1. Basis and Design Principles

The identification of a GenAI application competency framework for multimodal text writing is a top priority. The GenAI application competency framework for multimodal text

writing consists of criteria describing the level/requirements that students must achieve in multimodal text writing activities, as demonstrated by the development of GenAI application competencies in multimodal text writing. The process of developing the GenAI application competency framework for multimodal text writing includes establishing the concept of GenAI application competency in multimodal text writing, structuring GenAI application competency in multimodal text writing with specific components, describing those components, identifying behavioral indicators for each component, as well as behavioral quality criteria and specific levels for each criterion.

The development of the GenAI application competency framework in multimodal text writing is based on fundamental principles such as:

- The competency framework clearly demonstrates the ability to flexibly and purposefully apply the knowledge, skills, attitudes, needs, motivations, will, and interests of learners in multimodal text writing activities using GenAI as a support method, ensuring the development of multimodal text writing competencies and GenAI usage competencies in learning.
- The components of the competency framework must clearly display the requirements to be achieved in the multimodal text writing activities of the 2018 General Education Curriculum for the subject of Literature.
- Demonstrate the integration of skills related to information technology, critical thinking skills, professional skills (multimodal text writing), and moral principles in learning.

This study applies and builds upon existing competency frameworks for teaching, using information technology in teaching and education (such as DeSeCo by OECD, UNESCO, Stanford) [9], and adapts them to suit the characteristics of the domestic education system.

2.2.2. Structure of the Competency Framework

The competency to apply GenAI in writing multimodal texts includes the following basic dimensions:

Dimension 1: Foundational Understanding

Criterion 1.1: Understanding of GenAI

Criterion 1.2: Knowledge of multimodal text

Criterion 1.3: Basic understanding of the principles of using GenAI in learning

Dimension 2: Effective Interaction with GenAI

Criterion 2.1: Skill in prompt engineering

Criterion 2.2: Skills in evaluating the effectiveness of results generated by GenAI

Criterion 2.3: Skills in using appropriate GenAI tools

Dimension 3: Strategic Application in the Multimodal Writing Process

Criterion 3.1: Using GenAI in the preparation phase

Criterion 3.2: Using GenAI in the brainstorming and outlining phase

Criterion 3.3: Using GenAI in the drafting phase

Criterion 3.4: Using GenAI in the post-writing editing phase

2.2.3. Detailed Criteria, Indicators, and Competency Levels

The structure of GenAI application capabilities in multi-modal text writing is presented with specific components, descriptions of those components, identification of behavioral indicators for each component, as well as behavioral quality criteria and specific levels for each criterion (divided into three levels below: Novice, Developing, Proficient). Details are presented in Table 1 below.

Table 1. Detailed Competency Framework for Utilizing GenAI in Multimodal Text Writing

Competency Dimension	Criterion	Behavioral Indicator	Level 1: Novice	Level 2: Developing	Level 3: Proficient
1. Foundational Understanding	1.1. Understanding of GenAI	1.1.1. Understand the concept and know some popular GenAI tools	Understand the concept of GenAI.	Specifically present the concept of GenAI and list some GenAI tools	Explain and describe the concept of GenAI, know many GenAI tools
		1.1.2. Know the effects and limitations of GenAI	Know the effects but cannot list the limitations of GenAI	List some effects and limitations of GenAI	Clearly understand and explain the effects and limitations of GenAI
	1.2. Knowledge of multimodal text	1.2.1. Understand the concept of multimodal texts and identify the modalities in multimodal texts	State the concept of multimodal texts	Identify the concept and basic modes in multimodal texts	Understand and analyze the concept of multimodal text and modes in multimodal text
		1.2.2. Understand the role of multimodal text.	State the role of images in multimodal text	Identify the role of images, sound, and video in multimodal text	Clearly analyze the role of multimodal text in many cases
	1.3. Basic understanding of the principles of using GenAI in learning	1.3.1. Understand the ethical principles of using GenAI	Understand the principle of not copying GenAI	Understand the harm of GenAI abuse	Analyze the forms of violation when using GenAI
		1.3.2. Know how to evaluate discrepancies from GenAI	Know that information from GenAI may be incorrect	Recognize inaccurate information from GenAI	Analyze the causes and harms of inaccuracies from GenAI.
2. Effective Interaction with GenAI	2.1. Skill in prompt engineering	2.1.1. Write appropriate, effective prompts	Write concise prompts	Write basic prompts that meet requirements	Write detailed, specific prompts that produce appropriate results.
	2.2. Skills in evaluating the effectiveness of results generated by GenAI	2.2.1. Evaluate inaccurate results from GenAI	Identify incorrect information generated by GenAI	Compare with teacher-provided materials to identify unsuitable results from GenAI	Know how to clearly analyze inaccurate results generated by GenAI

Competency Dimension	Criterion	Behavioral Indicator	Level 1: Novice	Level 2: Developing	Level 3: Proficient
		2.2.2. Recognize results that are not suitable for the task generated by GenAI	Detect inaccurate words from GenAI	Fully indicate words and sentences that are not appropriate for the task generated by GenAI	Analyze the inappropriateness of the results generated by GenAI
	2.3. Skills in using appropriate GenAI tools	2.3.1. Select appropriate GenAI tools	Use familiar but inappropriate GenAI tools	Use a GenAI tool appropriate for the task	Select the most appropriate GenAI tools for each type of task required
3. Strategic Application in the Multimodal Writing Process	3.1. Using GenAI in the preparation phase	3.1.1. Using GenAI to find topics for writing	Requiring GenAI to suggest topics and ideas in a general way	Using GenAI to provide many suitable options	Using and evaluating the topics suggested by GenAI to find meaningful topics
		3.1.2. Using GenAI to search for writing materials	Ask GenAI to generate general materials	Ask GenAI to use the internet to search for authentic sources	Ask GenAI to use the internet to search for authentic, relevant, and effective sources
	3.2. Using GenAI in the brainstorming and outlining phase	3.2.1. Using GenAI to suggest ideas during brainstorming and outlining	Ask GenAI to brainstorm and outline without providing specific instructions	Provide specific instructions for GenAI to brainstorm and outline	Use GenAI as a tool to assist with brainstorming and outlining, then adjust as appropriate
	3.3. Using GenAI in the drafting phase	3.3.1. Using GenAI to edit written content	Copy GenAI's output into the article	Ask GenAI to edit written content but do not understand how to make the corrections	Analyze the errors GenAI corrected to adjust the article
		3.3.2. Use GenAI to suggest ideas for multi-modal modes	Use GenAI to draw diagrams and charts	Adjust the diagrams, charts, and images created by GenAI	Change and supplement the multi-modal modes created by GenAI
	3.4. Using GenAI in the post-writing editing phase	3.4.1. Use GenAI to correct spelling and basic expression errors	Request GenAI to automatically correct errors	Request GenAI to correct spelling and expression errors based	Use GenAI to proactively detect and correct errors, without relying entirely on GenAI

Competency Dimension	Criterion	Behavioral Indicator	Level 1: Novice	Level 2: Developing	Level 3: Proficient
				on specific suggestions	
		3.4.2. Use GenAI to cite reference materials	Use citations generated by GenAI	Use GenAI to generate citations and review them	Use GenAI to build and adjust citations to ensure they are appropriate for the research paper

From the GenAI application competency framework in multimodal text writing, teachers can develop a GenAI application competency development path in multimodal text writing. The GenAI application competency development path in multimodal text writing is understood as a description of the different levels of development of each competency that learners need or have achieved. The GenAI application competency development path in multimodal text writing may include the following levels: Level 1 - Not achieved; Level 2 - Achieved; Level 3 - Fair; Level 4 - Good. In the process of assessing students' GenAI application skills in multi-modal text writing, the GenAI application skill development path plays an important role as a standard for evaluating students' skills. It also reflects the development of each student's skills, showing where they are on the GenAI application skill development path in multi-modal text writing. Through the GenAI application competency development path in multi-modal text writing, students have a basis for self-assessing their GenAI application competency in multi-modal text writing to plan for honing the components of GenAI application competency in multi-modal text writing that they have not yet achieved. The GenAI application competency development path for writing multimodal texts also serves as a framework for students' GenAI application competency in writing multimodal texts, acting as a measure to determine the level of achievement in GenAI application competency in writing multimodal texts, thereby supporting the objective and accurate.

2.3. Proposing Solutions to Develop GenAI Application Capabilities in Multimodal Text Writing for Student

2.3.1. Equipping Foundational Knowledge to Develop GenAI Application Capabilities in Multimodal Text Writing for Students

Foundational knowledge refers to methodological knowledge that guides learners in solving problems and meeting the objectives of the subject of influence. Foundational knowledge is instrumental knowledge, serving a pivotal function in forming a foundation and cultural knowledge base for learners. While scientific knowledge represents systematic human understanding of natural and social laws and of oneself, instrumental knowledge focuses on the ability to apply such understanding in activities to continuously master the subject, meet set objectives, and develop and improve oneself. Foundational knowledge is highly condensed theory, which must ensure accuracy and scientific rigor while demonstrating applicability in various purposes, depending on the user.

To develop students' ability to apply GenAI in writing multimodal texts, teachers can focus on specific knowledge such as the following. First, foundational knowledge about multimodal texts is one of the essential types of foundational knowledge in developing students' ability to apply GenAI in writing multimodal texts. Basic knowledge of multimodal texts helps students identify the differences between monomodal and multimodal texts; understand the characteristics, roles, and classification methods of multimodal texts, as well as the basic elements that constitute modes in multimodal texts, thereby guiding them in the correct activities for writing this type of text. General knowledge about multimodal texts falls within the scope of theoretical literary

knowledge. This group of knowledge includes the following basic content: The concept of multimodal texts and modes in multimodal texts (such as language, images, diagrams, charts, data, symbols, sound, video, actions, links, etc.), how to identify multimodal texts, how to distinguish between multimodal and monomodal texts; different types of multimodal texts classified according to specific criteria. Each different criterion defines different sub-types. Second is basic knowledge about GenAI and how to use GenAI ethically in learning and research. Students need to identify the types of GenAI that are appropriate for writing multimodal texts, as well as the effects and limitations of GenAI when using it. On the other hand, ethical issues when using GenAI need to be addressed so that students know how to use GenAI effectively and do not abuse GenAI in the learning process.

The process of equipping students with background knowledge to develop GenAI application skills in writing multimodal texts is a long-term process that can be carried out in many different ways. Teachers can equip students with knowledge through diverse teaching methods in class, a system of homework assignments, or extracurricular activities. Among these, teachers can choose the flipped classroom model to implement the provision of foundational knowledge aimed at developing students' ability to apply GenAI in writing multimodal texts. The flipped classroom model is characterized by shifting basic knowledge acquisition activities from the classroom to self-study at home and increasing exchange and discussion activities on advanced topics in the classroom. This requires learners to not only mobilize but also actively apply their foundational knowledge during the teaching process.

2.3.2. Cultivate Effective and Appropriate Skills for Interacting with GenAI

The skill in prompt engineering and interact with GenAI is a fundamental skill for students working with GenAI. To prevent students from simply issuing prompts to GenAI and waiting for results, then using those results without critical thinking to adjust and evaluate, teachers need to train students' skills in interacting with GenAI. Instead of accepting all results generated by GenAI, students are required to critically evaluate the results, assess them to adjust and modify them to suit the task and appropriate contexts. To cultivate this skill, teachers can guide students on how to write an effective command that includes specific components such as objectives, context, role, desired format, tone, and other specific requirements. Based on this, teachers can provide a number of sample prompts for students to refer to. On the other hand, to help students think critically when giving prompts to GenAI, teachers need to guide students on how to ask follow-up questions when GenAI provides results. This encourages students to continue asking probing questions about the answers GenAI provides, rather than simply accepting them.

To practice this skill, teachers can have students write prompts through specific tasks in multimodal text writing. Based on practicing writing prompts like this, students can identify specific cases where GenAI should be used. Teachers can also edit the prompts students have written and learn from the process of students interacting with GenAI to ensure the most effective interaction.

When guiding students to interact with GenAI, teachers need to pay special attention to guiding students to interact in a two-way manner while ensuring that students use GenAI's results ethically. This is demonstrated by students being aware that the results generated by GenAI are only meant to guide their thinking and cannot replace their own thinking, while also understanding the harmful effects of overusing GenAI in learning, which can lead to students becoming lazy in their thinking, having poor reflexes, and losing creativity in their thinking and learning.

2.3.3. Applying AI Strategically in the Stages of the Multimodal Text Writing Process

This methodology aims to develop students' skills in applying GenAI to the practical writing of multimodal texts at each stage of the writing process: pre-writing; brainstorming and outlining; drafting; and revising. To implement this measure effectively, teachers need to develop students' thinking about using GenAI as a tool to support thinking rather than replacing students' thinking

or the actions they perform during the writing process. Teachers can implement this measure by guiding students to apply GenAI in each step of the writing process as follows:

- Pre-writing Stage:

During this stage, teachers can guide students in searching for and selecting topics or initial information to serve the process of writing multimodal texts. After using GenAI to suggest a number of suitable topics, students will proceed to select and evaluate the topics that have been suggested. GenAI can also help students find writing materials through verified internet search sources. Teachers need to guide students to select appropriate sources of information to obtain quality information for their writing while evaluating and eliminating inaccurate information.

Teachers can guide students in using GenAI to generate and develop key ideas. As GenAI presents these ideas, teachers should guide students rigorously assess and select, adjusting the content written by others to create a suitable outline, avoiding bias.

- Writing Stage:

Teachers can guide students in using GenAI to assist with drafting. In this case, GenAI plays a role in adjusting the content students write and providing options for students to improve their written work. Teachers should guide students on the need to avoid directly using specific ideas and sentences generated by GenAI.

In addition, GenAI can also support the creation of multimodal elements for students. Teachers should note that if images or audio created by GenAI are used, they must clearly annotate this information to ensure scientific ethics.

- Revising Stage:

Teachers can guide or provide students with suggestions for developing checklists or rubrics to assess the quality of their writing. At the same time, students can also use GenAI as a tool to search for and detect errors in expression, spelling, word usage, sentence structure, and grammar. Based on this, students can self-correct the errors detected by GenAI. Note that students must still read their work and self-correct, writing their own reflections after completing the essay. On the other hand, when receiving feedback from GenAI, students also need to provide appropriate counterarguments, as not all errors identified by GenAI are accurate.

In summary, teachers need to guide students in using GenAI tools appropriately at each stage. The important thing is not to let GenAI write for students, but rather to guide their thinking, learning, and development of creativity.

3. Conclusion

This study has highlighted the urgent need to develop GenAI application capabilities in teaching multimodal text writing to secondary school students in order to meet the requirements of the 2018 general education curriculum. Based on the development of a competency framework for GenAI application in students' multimodal text writing, the study proposed appropriate pedagogical solutions to ensure the use of GenAI as a suitable and effective support tool in the process of students writing multimodal texts. This study is an important suggestion for teachers to guide students on how to apply GenAI in writing multimodal texts, aiming to develop digital literacy while also developing students' ability to create multimodal texts, meeting the increasingly high demands of educational reform.

REFERENCES

- [1] Discourse Analyzer, (2025). Kress and van Leeuwen in discourse analysis. Retrieved April 20, 2025, from <https://discourseanalyzer.com/kress-and-van-leeuwen-in-discourse-analysis/>

- [2] Ministry of Education and Training, (2018). *General Education Curriculum for Language Arts*, (Issued in conjunction with Circular No. 32/2018/TT-BGDĐT dated December 26, 2018, by the Minister of Education and Training) (Vietnamese).
- [3] Nguyen TX, (2023). Managing the teaching of Literature under the 2018 general education curriculum. *Journal of Educational Equipment: Education Management*, 1(284) (Vietnamese).
- [4] Chanpradit T, (2025). Generative artificial intelligence in academic writing in higher education: A systematic review. *Edelweiss Applied Science and Technology*, 9, 889-906.
- [5] University of Rochester, (n.d.). Generative AI use in education. Retrieved April 22, 2025, from <https://www.rochester.edu/provost/gen-ai-education/>
- [6] Kress G & van Leeuwen T, (2001). *Multimodal discourse: The modes and media of contemporary communication*. London: Arnold.
- [7] Tran TN, (2020). Some methods for teaching reading comprehension of multimodal texts in Literature. *Vietnam Journal of Educational Sciences* (31) (Vietnamese).
- [8] Serafini F, (2014), *Reading the Visual: An Introduction to Teaching Multimodal Literacy*, Teachers College Press, New York.
- [9] DeSeCo, (2016). *Definition and selection of competencies: Theoretical and conceptual foundations*. Retrieved April 12, 2025, from <https://www.deseco.ch/bfs/deseco/en/index.html>