

The Potential and Implications of AI-Generated Feedback for Primary School Composition Writing

Qizhong Chang¹, Zoey Chow²

^{1,2}National Institute of Education, Singapore

Emails: ¹qizhong.chang@nie.edu.sg, ²chow_zi_hui_zoey@moe.edu.sg

Abstract

Effective writing instruction requires regular and detailed feedback; however, it is not always easy to provide effective feedback and the provision of such feedback can be taxing for teachers (Bai et al., 2013). Singapore's Ministry of Education (MOE) has leveraged Artificial Intelligence (AI) – in the form of automated writing evaluation (AWE) systems in language classrooms – to provide basic feedback on students' writing in areas such as spelling and grammar which will allow teachers to focus on “guiding students in the more complex aspects of language construction, and to develop higher-level skills like creative expression, persuasiveness and tone” (MOE, 2023). In this paper, we are taking a critical and evaluative look in a richly contextualised setting at whether Generative AI (GAI) platforms, such as Gemini, can indeed fill the gap to provide feedback on higher-order thinking skills such as critical thinking, creativity, and organisation (Gupta et al., 2024). We address two research questions: (1) To what extent can GAI provide effective feedback on continuous writing tasks for primary school students in Singapore? (2) What are the potential implications of GAI-generated feedback on writing instruction? In our analysis of feedback patterns, GAI consistently focuses on content and style, suggesting improvements such as richer vocabulary and stronger conclusions, without actually providing clear examples. GAI also exhibits a lack of personalisation, not using personal pronouns or highlighting student strengths. Additional prompts from teachers are required to elicit more granular feedback and a more positive tone. Lastly, while GAI can assess essays based on a given rubric, it struggles with evaluating narrative structure, requiring further teacher guidance. Overall, GAI has the potential to be a valuable tool in language and writing classrooms, offering a starting point for feedback that requires both teacher expertise and teachers' digital literacy to personalise and refine.

Keywords: Feedback, Generative Artificial Intelligence, Continuous Writing, Writing Instruction

Introduction

Assessment plays a crucial role in teaching as it serves to provide information that is crucial to improving learning (Wong et. al., 2020). In the past decade, there has been a shift in the education landscape in Singapore as there have been major national educational initiatives implemented that place a strong emphasis on Assessment for Learning (AfL) and Assessment of Learning (AoL), which refer to formative and summative assessment respectively (Wong et. al., 2020).

Feedback is a crucial part of AfL as it allows for learning gaps to be closed. Feedback is “information provided by an agent...regarding aspects of one's performance” (Hattie &

Timperley, 2007, p.81), and ‘a process through which learners make sense of information from various sources and use it to enhance their work’ (Carless & Boud, 2018, p.1315). Effective writing instruction requires regular and detailed feedback; however, it is not always easy to provide effective feedback due to gaps in writing teachers’ feedback literacy (Lee, 2021), and the feedback cycle might require a lot of time on the part of educators (Bai et. al., 2013).

With Artificial Intelligence (AI) quickly establishing itself as a crucial and transformative force in various industries in the past year, AI in Education (AIED) has been the focus of many studies (Zhang & Aslan, 2021). Generative Artificial Intelligence (GAI) is a subset of AI that focuses on generating novel content such as text, code, or creative content forms. One example of GAI is Gemini (formerly Bard), a system that demonstrates advanced capabilities in natural language processing, allowing it to analyze text, identify strengths and weaknesses, and offer feedback on various aspects of writing, including grammar, vocabulary, and sentence structure (Gemini Team, 2023). GAI holds much promise for assisting teachers in various tasks, including providing automated feedback on student writing (Gupta et. al., 2024).

Recent systematic reviews of AIED have shown an upward trend in both regional and global adoption of AI applications in primary education. These applications span a wide spectrum, encompassing those for adaptive learning and personalised tutoring, intelligent assessment and management, profiling and prediction (Wang et al., 2024). AI is also prominently featured in the EdTech Masterplan 2030 by Singapore’s Ministry of Education (MOE). The plan has leveraged new technologies by launching AI features in the Singapore Student Learning Space (SLS), an online learning portal used by schools, which is meant to “customise students’ learning experiences and augment the guidance given by teachers” (MOE, 2023). Two Learning Feedback Assistants - Language Feedback Assistant for English (LangFA-EL) and the Short Answer Feedback Assistant (ShortAnsFA) - were launched in December 2023. The LangFA-EL can provide basic feedback on students’ writing in areas such as spelling and grammar which will allow teachers to focus on “guiding students in the more complex aspects of language construction, and to develop higher-level skills like creative expression, persuasiveness and tone” (MOE, 2023). On the other hand, the ShortAnsFA focuses on close-ended and short-answer questions. It can provide suggested grades and auto-generated content-related feedback which allows teachers to shorten feedback time by using the AI-generated feedback and grades as a draft that they can edit and build on. At this time, there are no studies detailing the performance and shortfalls of LangFA-EL or ShortAnsFA. However, MOE (2023) issues the disclaimers “As LangFA-EL and ShortAnsFA uses generative AI, its response is probabilistic in nature and this results in it being inaccurate at times...Bear in mind that LangFA-EL and ShortAnsFA may have also missed some errors”. Therefore, it remains an open question if AI can actually produce quality feedback that is useful and effective for teaching and learning.

Research Gap

While the potential of GAI for automated feedback in writing is gaining traction, significant research gaps remain, particularly regarding its efficacy and implementation in specific educational contexts. The use of GAI is a relatively recent development; thus, limited research exists on the use of GAI for feedback in the Singaporean education system, especially within the specific context of primary school composition writing, which often emphasizes specific content and language features aligned with the Singapore curriculum. Current research on GAI for writing feedback primarily focuses on higher education settings or general-purpose writing tasks. These studies often highlight the limitations of GAI systems in providing feedback on

higher-order thinking skills such as critical thinking, creativity, and organization (Gupta et. al., 2024).

This study aims to address the research gap by focusing on the potential of GAI, specifically Gemini, in providing feedback on continuous writing for Singaporean primary school students. It is important here to note that while the MOE feedback systems do allow for basic feedback focusing on form (grammar, spelling) in short-answer questions, this study aims to investigate the possibility of using GAI to provide feedback on higher-level skills like creativity, stylistic choices and content.

Research Questions

This study addresses two primary research questions: (1) To what extent can GAI provide effective feedback on continuous writing tasks for primary school students in Singapore? and (2) What are the potential implications of GAI-generated feedback on the teaching and learning process in Singaporean primary school composition writing?

The first question delves into the effectiveness and accuracy of Gemini's feedback. It will investigate whether the feedback provided by Gemini aligns with established grading rubrics used in Singaporean primary schools and whether it can offer constructive suggestions and effective feedback for improvement. This will also address issues of writing teachers' feedback literacy, and the time pressures in the feedback cycle.

The second question focuses on the wider impact of GAI-generated feedback on the learning process and student engagement. It will explore how GAI feedback might influence students' understanding of writing concepts, their approaches to revising their writing and self-reflection, and their overall motivation and engagement with writing tasks. This question also investigates potential opportunities and challenges associated with using GAI feedback, such as fostering self-directed learning, promoting deeper comprehension, and addressing potential concerns related to overdependence or reliance on technology and ethical concerns about AIED.

This research has the potential to contribute to the ongoing dialogue on the ethical and pedagogical implications of utilizing AI in education. By addressing these research questions, this study aims to provide valuable information for educators and policymakers in considering the integration of GAI for feedback in classrooms, while also highlighting considerations for maximizing its effectiveness and student learning outcomes.

After analyzing the data, a few main recurring patterns were found. A key takeaway from analyzing Gemini's feedback patterns is that it offers consistency in certain areas while requiring teacher intervention in others. Consistently, Gemini focuses on content and style, suggesting improvements like richer vocabulary and stronger conclusions. It also identifies specific action points for improvement but may lack clear examples. To address this, teachers might need to provide additional prompts to elicit more granular feedback from Gemini. Another consistent feature is Gemini's lack of personalization. It doesn't use personal pronouns or highlight student strengths. Educators can address this by prompting Gemini for a more positive tone or by personalizing the feedback themselves to manage students' emotional receptiveness. Finally, while Gemini can assess essays based on a rubric, it seems to have limitations in addressing overall narrative structure. Teachers may need to provide additional guidance in this area. Overall, Gemini appears to be a valuable tool for educators, offering a starting point for feedback that requires teacher expertise to personalize and refine.

Literature Review

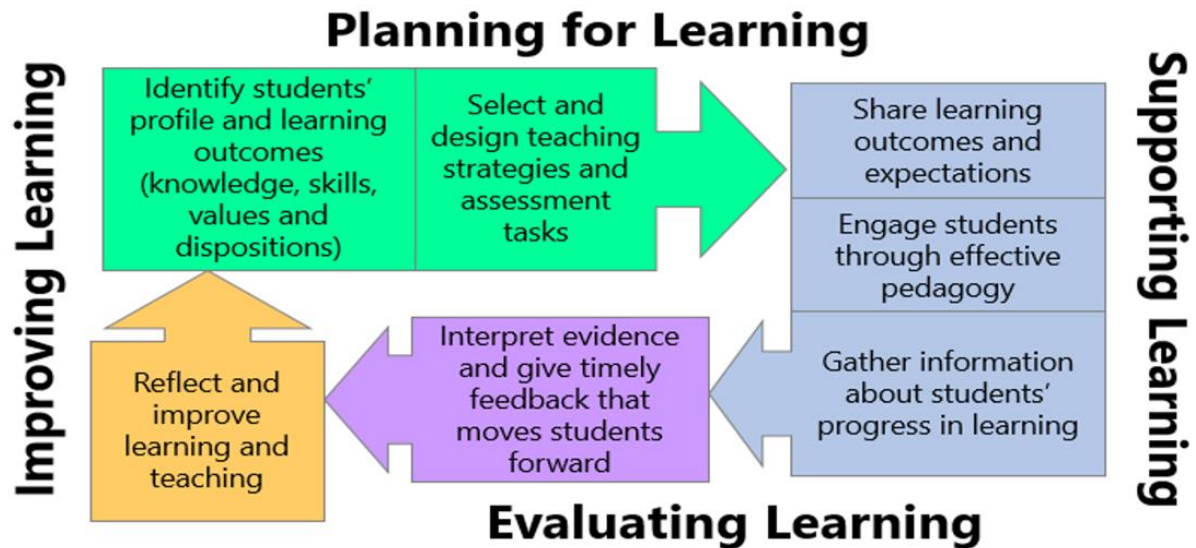
Feedback Types

Feedback in an education/writing context is traditionally classified into several categories based on its purpose and nature. Two important distinctions are formative vs. summative feedback and direct vs. indirect feedback. Direct feedback provides explicit corrections or suggestions, while in indirect feedback, the instructor indicates that an error exists but does not provide the correction (Lalande, 1982). Formative feedback focuses on improving the learner's skills and understanding during the writing process, and is iterative. In contrast, summative feedback evaluates the final product, often in the form of grades or comprehensive comments summarising the strengths and weaknesses of the piece (Bloom et. al., 1971). AI-generated feedback can theoretically fit in all these established categories, especially when given explicit prompts to act in certain specific ways. For instance, while many AI writing tools such as Grammarly and Gemini inherently provide direct feedback (such as suggesting corrections for grammar or word choice), they can be given instructions to only offer indirect feedback (such as identifying areas that lack clarity or coherence) without explicitly dictating how to fix them. AI tools are naturally compatible with formative feedback, as it can provide continuous and iterative input during the drafting and writing process. AI systems can also generative summative feedback, but this is less common and has to be based on predefined rubrics or criteria.

Assessment for Learning (AfL)

There has been extensive research done that validates the efficacy of AfL practices and shows that there is a possibility of significant learning improvement when AfL is part of teachers' classroom practices (Black & Wiliam, 1998; Brookhart, Moss, & Long, 2009; Carless, 2012; Heritage, 2007; Leahy & Wiliam, 2012; Natriello, 1987; Wiliam, 2011; Wiliam & Leahy, 2015, as cited in ELIS, 2018). In Singapore, the implementation of AfL was based on the work of Black and William (1998) in which the central tenet focused on four key activities — sharing success criteria with learners, classroom questioning, comment-only marking and peer- and self-assessment (ELIS, 2018). From that, three key processes for feedback were highlighted: (1) where the learners are in their learning, (2) where they are going and (3) what needs to be done to get them there. MOE (2013) has also introduced the concept of an assessment-feedback learning cycle as seen in Figure 1 below.

Figure 1. MOE Assessment-Feedback Learning Cycle



As part of the Singapore Teaching Practice (STP), MOE also encourages the use of the teaching action, “Comment only feedback” (STP, 2024). This teaching action has teachers providing comments that are substantial and clear enough to guide students in improving their work without depriving them of the chance to do the work or to think about the solution themselves. In order to achieve this, it was suggested that comments can be phrased as questions which can aid in scaffolding students' thinking to take ownership of their own learning. It states that there should not be feedback at the self-level as it is ineffective because it focuses on personal evaluations and affects. For example, compliments like “well done” or “good job” simply do not provide any information on how students can improve their performance on the given task and can direct attention away from the task, process or self-regulation levels. Additionally, the STP discourages the use of judgmental feedback with words that imply blame, fault, mistake, and incompetence as teachers should seek to build a positive classroom culture that can encourage students' participation and co-construction (STP, 2024).

Nicol & Macfarlane-Dicks's Principles of Feedback (2006)

Nicol and Macfarlane-Dick (2006) argue for a shift towards student-centered feedback practices in educational settings. They contend that traditional approaches, where teachers solely “transmit” feedback to students, are insufficient to promote self-regulated learning, a crucial skill for lifelong learning. Their model emphasizes the importance of students actively interpreting, responding to, and generating feedback, including self-assessment and feedback from peers and teachers. This student-centered approach acknowledges the cognitive, motivational, and behavioural aspects of learning. Effective feedback, according to the authors, should address all three aspects to facilitate meaningful learning.

Nicol and Macfarlane-Dick (2006) propose seven principles for effective feedback, each one addressing the cognitive, motivational and behavioural aspects of self-regulation:

1. **Clarity of Goals and Criteria:** Providing clear examples, criteria sheets, and opportunities for discussion about expectations helps students understand what good performance looks like.
2. **Facilitating Self-Assessment:** Encouraging students to self-assess their work, either individually or with peers, promotes reflection and ownership of learning.

3. **High-Quality Information:** Feedback should be timely, specific, and actionable, focusing on improvement rather than solely highlighting strengths and weaknesses.
4. **Encouraging Dialogue:** Peer discussions and classroom conversations about feedback can enhance understanding and solidify learning.
5. **Positive Motivation and Self-Esteem:** Feedback should encourage effort and progress, emphasizing the learning process over grades.
6. **Closing the Performance Gap:** Providing opportunities for resubmission, retaking quizzes, and identifying action points helps students bridge the gap between their current and desired performance.
7. **Informing Teaching:** Effective feedback practices provide teachers with valuable data to adapt and improve instruction.

These principles can also be seen in the key characteristics of effective feedback that are highlighted in ELIS (2018):

1. Feedback should give the learner information about the assignment or the process of learning that bridges the understanding gap between what is understood and what is intended to be understood.
2. Feedback needs to be precise, concise, and linked to the learning objectives.
3. Feedback needs to offer recommendations and hints on how to do better.
4. Feedback should be clear and simple, appropriate for the learner's cognitive level.
5. Feedback needs to be provided promptly, for instance, upon the completion of a new assignment or, in the case of learners making slow progress, throughout the acquisition of new abilities and ideas. Delayed feedback could be a better option for more complex assignments requiring a lot of processing time.

Studies emphasize the importance of positive teacher language in fostering student engagement and language acquisition (Sun, 2021; Ni, 2012). Positive teacher feedback, including praise and care, can create a supportive learning environment that lowers students' affective filter, the emotional barrier that can hinder language learning (Krashen, 1982). This allows students to be more receptive to feedback and instruction. Therefore, incorporating positive language into feedback strategies is crucial for maximizing student learning.

Effective writing instruction in primary schools is crucial for fostering students' literacy development and communication skills (Sucipto, 2014). Additionally, research suggests that well-designed feedback can promote student learning and improve writing outcomes and development. However, providing such comprehensive and individualized feedback can be a significant challenge for educators, particularly when dealing with large class sizes and diverse student needs (Xu and Harfitt, 2018).

GAI and Writing Feedback

Recent advancements in AI have paved the way for exploring the potential of AI tools in providing automated feedback on student writing. These tools can analyze student work based on predefined parameters or learn from human-annotated data to identify strengths and

weaknesses (Liu et. al., 2023). According to Liu et. al. (2023), AI-powered tools like Gemini can potentially:

1. Automate initial feedback: AI can analyze student writing and provide initial feedback on aspects like grammar, mechanics, and vocabulary usage, freeing up educators' time for more personalized guidance.
2. Offer personalised insights: AI tools can be trained on vast datasets of student writing, allowing them to provide feedback tailored to individual student needs and writing styles.
3. Scale-up feedback provision: AI can efficiently analyze large volumes of student writing, potentially addressing the challenges of providing timely feedback in large classrooms.

While AI offers promise in terms of efficiency and scalability, limitations exist. Current AI models might struggle to capture the full nuances of human language, potentially overlooking aspects like emotional depth, stylistic ingenuity, or cultural context (Oritsegbemi, 2023). Additionally, ethical considerations regarding potential biases within AI models and student data privacy must be carefully addressed (Slimi and Carballido, 2023). AI writing tools can unintentionally perpetuate biases, impacting marginalised students through language and cultural bias, and unfair assessment practices. Fluency in standard registers and varieties of English is often assumed, disadvantaging multilingual students or those with distinct dialects (Herold, 2022). GAI used for evaluating student work may reinforce these biases, leading to achievement gaps for underrepresented students (Mhlanga, 2023). Over time, such biases can become entrenched, further exacerbating inequalities in education (Selwyn, 2019).

In second language (L2) writing classrooms, the introduction of GAI-assisted tools has also led to a sea change in the conceptualisation and practice of automated written corrective feedback (AWCF) in L2 pedagogy. In a mixed-method multiple case study with four L2 writers, Yan & Zhang (2024) found that behavioural engagement with ChatGPT as a provider of AWCF was influenced by individual differences in language proficiencies and technological competencies. Participants in the study struggled to effectively regulate the learning process metacognitively. While the tool created an affectively engaging environment, it was also competence-demanding and time-consuming for learners.

AI in education can lessen the workload of teachers, enhance marking uniformity, offer widely customized learning, and guarantee more consistency in the instruction offered by schools and other educational institutions (Fischer, 2023). Thus, there may be cases of both augmented marking (where people and machines mark humans together) and automated marking (where robots mark humans).

Research on Model Essays in Language Learning

Studies have shown that analyzing model essays can improve students' understanding of writing expectations, vocabulary usage, and sentence structure (Abe, 2008). This aligns with the Singaporean curriculum, which emphasizes specific content and language features in primary school writing (CPDD, 2020). It's common for parents in Singapore to purchase composition books for their kids so they can practice replicating sample compositions to become more proficient in writing (Gupta, 1995). In order to finish their writing tasks or to review their writing techniques before tests, a lot of elementary school students study model compositions. Additionally, they receive a list of vocabulary words to learn and include in their writings.

According to Abe (2008), model essays can serve as:

1. Points of reference: Students can analyze the structure, language use, and content organization within model essays to gain insights into effective writing strategies
2. Self-assessment tools: By comparing their own writing to model essays, students can identify areas for improvement and develop their critical evaluation skills.

Methodology

This study explores the potential of utilizing narrative writing models, predetermined rubrics and AI feedback to enhance primary school students' writing skills. The methodology focuses on the analysis of pre-existing model essays and subsequent AI-driven feedback to identify potential benefits for classroom practice.

Data Collection

Selecting Model Essays

The study employed a convenience sampling approach, selecting readily available resources for analysis. Model essays were specifically chosen from four assessment books catering to primary school students. These assessment books typically showcase examples and guides for various writing prompts. The assessment books chosen were readily available and found in common bookstore chains in Singapore. Additionally, they were chosen as at least two of the books contained similar themes and topics of model essays.

The rationale behind selecting model essays stems from their potential to act as feedback tools. Model essays, often written by native or near-native speakers, can offer valuable insights regarding effective structure, engaging vocabulary choices, and compelling content. The choice to analyse model essays shifts the focus of feedback from simply correcting grammatical errors (lower-order skills) to a deeper analysis of content, language use, and structure (higher-order skills). The various language and phrasing errors typical in many student essays will also potentially affect GAI's ability to clearly understand plot details and structure, and provide good feedback. Another practical reason for selecting model essays over student-written ones is the fact that there is difficulty in collecting high-quality student-written essays from schools in Singapore, due to access reasons.

The use of model essays presents a practical implication for classroom practice. Students can initially refer to the model essays, then receive AI-generated feedback tailored to their own writing. This combined approach can raise awareness of essential aspects to consider during the writing process, encouraging students to focus on more than just grammar and sentence structure.

Selecting Prompts and Analyzing Models

Five common writing themes were chosen for analysis: A Prank, Celebrations (Chinese New Year), The Bully, A Jovial Person, and Littering. This selection reflects themes that are regularly used in primary school assessments and writing exercises.

For each prompt, two distinct model essays were sourced from different assessment books (see Appendix A). This approach aimed to capture the diversity of writing styles and approaches within the chosen themes. This would also allow for the identification of common elements or

patterns between the two model essays to see if AI feedback would provide common feedback despite the content being different.

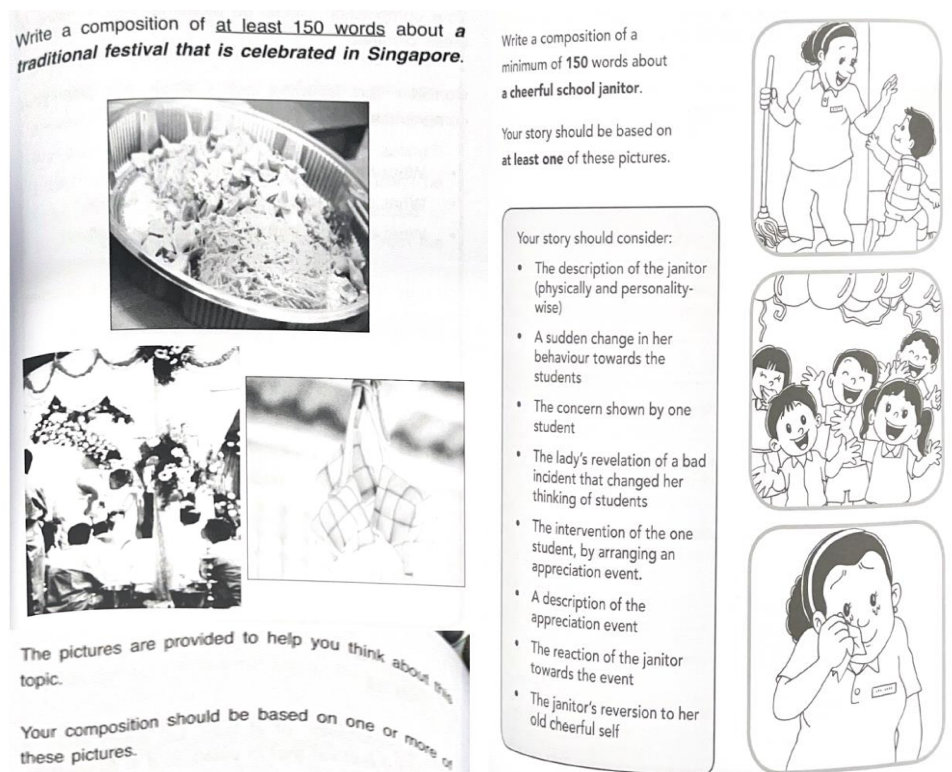
Of note is that two of the essays selected, “Littering” and “A Jovial Person” do not follow the typical narrative structure — the essays have no clear beginning, middle and end. Instead, the two compositions seem to be focused on a specific character’s thoughts, experiences and feelings about the given topic. These essays were chosen to see if Gemini would focus on giving feedback related to narrative story structure.

AI Tool Selection and Rationale

Gemini version 1.0 (previously known as Google Bard) was chosen as the primary tool for AI feedback generation. This choice stemmed from the limitations of the free version of ChatGPT (GPT-3.5) in not being able to accept image prompts. While ChatGPT offers paid subscription plans with image prompting capabilities, the free version of Gemini allows for the integration of images alongside written prompts.

In Singapore, primary school continuous writing tasks typically provide students with three images to aid them in their writing. The instructions for the continuous writing task typically require students to use at least one of the three images provided in their writing. The images given can be symbolic (i.e., an image of a Chinese New Year decoration), though they typically show characters in certain situations (i.e., characters celebrating something). Two examples of the instructions given to students for the writing task can be seen in Figure 2 below.

Figure 2. Examples of Continuous Writing Instruction (Sources: Casco Publications, 2014, pp. 149-151 and Raffles Publications, n.d., pp. 182-184)



With the affordance of being able to upload images to Gemini, it is assumed that the image prompts may further enhance the context and understanding of the writing prompts for the AI model, leading to more relevant and targeted feedback.

While this study utilizes Gemini for its specific functions, the broader concept of utilizing AI-powered feedback alongside model essays remains applicable across various platforms and tools. That is to say, the study should be replicable on other AI systems based on large language models, such as GPT-4 and Llama.

Utilizing a Pre-existing Rubric for Scoring

The study employed a pre-existing rubric that was taken from a school in Singapore (See Figure 3), commonly used in primary school settings, to establish a framework for assessing the language and content of student essays. The chosen rubric was cross-checked with official syllabus documents and those from other schools to ensure both consistency and authenticity. The rubric encompasses key criteria such as grammar, vocabulary, sentence structure, organization, and clarity of ideas, aligned with typical benchmarks for primary school writing assessment.

Figure 3. Rubric for Composition Writing**Language Marking Scheme for Primary 5 and 6**

Mark range	1 - 5	6 - 10	11 - 14	15 - 17	18 - 20
	Evident lack of organisation of events, numerous grammatical, punctuation, and spelling errors, and poor vocabulary usage.	Multiple errors in grammar, punctuation and spelling, Wrong and limited use of vocabulary, mediocre organisation and sequencing of events.	A number of errors in punctuation and spelling; Average range of vocabulary, wrong usage of words, average organisation and sequencing of events.	Only a few errors in grammar, punctuation and spelling; Impressive vocabulary usage and clear organisation and sequencing of events.	Very few minor errors in grammar, punctuation and spelling; Excellent range of vocabulary; Exceptional organisation and sequencing of events.

Content Marking Scheme for Primary 5 and 6

Mark range	1 - 5	6 - 10	11 - 14	15 - 17	18 - 20
	Confused, illogical, and totally irrelevant to the topic.	Ideas are somewhat vague, slightly confusing, and slightly relevant to the topic.	Clear, logical, and relevant ideas.	Interesting clear, logical and relevant ideas with a well-developed story.	Highly interesting, clear, logical and totally relevant ideas that help develop a very interesting story.

Steps for Generating AI Feedback

In order to collect feedback, the first step was to upload the rubric with the prompt, “This is a marking rubric for continuous writing for Primary 5 students in Singapore” (See Figure 4).

Figure 4. Input into Gemini and Response about Rubric

Z This is a marking rubric for continuous writing for Primary 5 students in Singapore.

Sample Marking Scheme for Primary 5 and 6

Mark range	1-5	6-10	11-14	15-20
Content	Ideas are confused, illogical, and irrelevant to the topic.	Ideas are somewhat vague, slightly confusing, and slightly relevant to the topic.	Ideas are clear, logical, and relevant.	Ideas are interesting, clear, logical, and relevant, with a well-developed story.
Language	Evident lack of organisation of events, numerous grammatical, punctuation, and spelling errors, and poor vocabulary usage.	Multiple errors in grammar, punctuation and spelling, wrong usage of words, average range of vocabulary, mediocre organisation and sequencing of events.	Only a few errors in grammar, punctuation and spelling, impressive vocabulary usage and clear organisation and sequencing of events.	Very few minor errors in grammar, punctuation and spelling, average range of vocabulary, excellent organisation and sequencing of events.

Standard Marking Scheme for Primary 5 and 6

Mark range	1-5	6-10	11-14	15-20
Content	Ideas are confused, illogical, and irrelevant to the topic.	Ideas are somewhat vague, slightly confusing, and slightly relevant to the topic.	Ideas are clear, logical, and relevant.	Ideas are interesting, clear, logical, and relevant, with a well-developed story.
Language	Evident lack of organisation of events, numerous grammatical, punctuation, and spelling errors, and poor vocabulary usage.	Multiple errors in grammar, punctuation and spelling, wrong usage of words, average range of vocabulary, mediocre organisation and sequencing of events.	Only a few errors in grammar, punctuation and spelling, impressive vocabulary usage and clear organisation and sequencing of events.	Very few minor errors in grammar, punctuation and spelling, average range of vocabulary, excellent organisation and sequencing of events.

Show drafts

✦ The image you sent is indeed a marking rubric for continuous writing for Primary 5 and 6 students in Singapore. It outlines the expectations for both content and language use, and assigns marks accordingly.

Content

- **1-5 marks:** Ideas are confused, illogical, and irrelevant to the topic.
- **6-10 marks:** Ideas are somewhat vague, slightly confusing, and slightly relevant to the topic.
- **11-14 marks:** Ideas are clear, logical, and relevant.
- **15-17 marks:** Ideas are interesting, clear, logical, and relevant, with a well-developed story.
- **18-20 marks:** Ideas are highly interesting, clear, logical, and totally relevant, that help develop a very interesting story.

Language

- **1-5 marks:** Evident lack of organisation of events, numerous grammatical, punctuation, and spelling errors, and poor vocabulary usage.
- **6-10 marks:** Multiple errors in grammar, punctuation and spelling, wrong usage of words, average range of vocabulary, mediocre organisation and sequencing of events.
- **11-14 marks:** Only a few errors in grammar, punctuation and spelling, impressive vocabulary usage and clear organisation and sequencing of events.
- **15-17 marks:** Very few minor errors in grammar, punctuation and spelling, average range of vocabulary, excellent organisation and sequencing of events.
- **18-20 marks:** No errors in grammar, punctuation and spelling, exceptional range of vocabulary, exceptional organisation and sequencing of events.

This marking rubric can be a helpful tool for both teachers and students to understand the expectations for continuous writing in Primary 5 and 6 in Singapore. It can help students to focus on developing their ideas, using a variety of vocabulary and grammar correctly, and organizing their writing in a clear and logical way.

Next, the essays are uploaded individually. The relevant images associated with the writing prompt were uploaded to Gemini. Each composition was uploaded as a new conversation so

that the system would not be influenced/trained by the user's previous inputs. According to Team Gemini (2023), while Gemini cannot directly learn from prior interactions, it can leverage context and its internal memory to tailor its responses to the specific composition. This new conversation format helps mitigate the influence of past interactions, allowing us to see if Gemini can offer effective feedback without user-specific training.

Using the uploaded rubric and images, Gemini was instructed to perform two specific tasks (See Figure 5). Firstly, it would analyze the provided essay's language and content based on the criteria outlined in the uploaded rubric. This analysis aimed to generate specific scores (not ranges) for each rubric category. Secondly, Gemini was asked to offer detailed feedback on the essay's language and content, addressing areas of strength and areas for improvement.

Figure 5. Example of Prompt Input into Gemini

Using the above marking rubric for both language and content, refer to these pictures and the writing prompt, "Littering"

1. Mark/score the language and content of the essay with specific marks (not a range) and a reason for the score
2. Give specific and detailed feedback on the following essay to make it better:

While providing useful initial feedback (See Figure 6), Gemini's analysis required further prompting to elicit deeper insights.

Figure 6. Example of feedback output by Gemini



Scoring and Feedback on Essay: "Littering"

Language Score: 18/20

- **Strengths:**

- Clear and concise sentences with proper grammar and punctuation.
- Varied vocabulary demonstrating good word choice.
- Descriptive use of language to paint a vivid picture of the scene (e.g., "scorching sun shone its rays", "dampened gnarly fingers").
- Use of transition words and phrases to connect ideas smoothly.

- **Areas for improvement:**

- Minor instance of word choice: "reeling back in shock" is slightly redundant as "stunned" already conveys the effect.

Content Score: 18/20

- **Strengths:**

- Addresses the prompt "Littering" by showcasing the negative consequences of littering through the incident.
- Presents a clear cause-and-effect relationship between the littering and the potential harm it could have caused.
- Highlights the positive aspects of helpfulness and community through Gopal's actions and the police response.
- Offers a clear and concise conclusion that reinforces the message of safety and responsibility.

- **Areas for improvement:**

- The essay primarily focuses on the specific incident and the resolution. While impactful, a broader discussion about the general effects of littering on the environment and society could further strengthen the content.

Overall Feedback:

This essay demonstrates strong writing skills with clear language, vivid descriptions, and a well-structured narrative. It effectively addresses the prompt by showcasing the dangers of littering and the importance of community responsibility. For further improvement, consider incorporating a broader discussion on the general impacts of littering beyond the specific incident. This will add depth to the content and create a more comprehensive response to the prompt.

Additional specific questions were asked, such as: "What specific steps could the student take to achieve a full score for language?" (See Figure 7) and "What specific steps could the student

take to achieve a full score for content?" (See Figure 8). This encouraged Gemini to provide actionable suggestions as well as a section of a model essay.

Figure 7. Gemini Output after Further Prompting with the Question,

“What specific steps could the student take to achieve a full score for language?”

- ✦ While the essay has strong language overall, achieving a perfect score in this section requires meticulous attention to detail and maximizing creative expression. Here are some specific steps the student could take to aim for a full score in language:

1. Enhance Figurative Language:

- While the essay uses descriptive language effectively (e.g., "scorching sun shone its rays"), consider incorporating additional figures of speech like metaphors, similes, or personification to add depth and imagery. For example, instead of "scorching sun," the writer could say "the sun beat down mercilessly, like a hammer on an anvil."
- Use sensory details beyond sight. Can you describe the sounds of the city, the smell of the air after the flower pot fell, or the texture of Mrs. Becca's bag? Engaging multiple senses can create a more immersive reading experience.

2. Refine Sentence Structure:

- While the sentences are grammatically correct, consider incorporating a variety of sentence structures to avoid monotony. This could involve using a mix of short and long sentences, introducing complex sentence structures with dependent clauses, or varying the sentence beginnings.
- Employ transitional words and phrases strategically to create smooth transitions between ideas and enhance the flow of the narrative.

3. Explore Vocabulary:

- Though the vocabulary is varied, consider using even more precise and evocative words to elevate the language further. Instead of "grateful," perhaps "overjoyed" or "heartfelt" could better capture Mrs. Becca's emotions. This requires thesauruses and careful selection of words that fit the context and tone.

4. Proofread and Edit:

- While the essay seems grammatically sound, meticulously proofread for any typos, punctuation errors, or minor inconsistencies in word choice. This final polish ensures the language is clear, concise, and error-free.

Remember, achieving a perfect score requires exceptional execution and often comes after multiple revisions. By incorporating these steps and focusing on refining the language, the student can significantly elevate the essay's literary quality and strive for a full score in the language category.

Figure 8. Gemini Output after Further Prompting with the Question,

“What specific steps could the student take to achieve a full score for content?”



While the essay already achieved a high score in content, here are some specific steps the student could take to aim for a perfect score:

1. Deepen the exploration of the characters:

- **Gopal:** Go beyond depicting Gopal's helpfulness. Briefly describe his initial reaction to seeing Mrs. Becca struggle, his thoughts and emotions during the incident, and his sense of accomplishment after the situation is resolved.
- **Mrs. Becca:** Briefly touch upon her feelings of vulnerability and fear during the incident. Perhaps mention her thoughts about the culprit or her appreciation for Gopal's assistance.

2. Expand on the resolution:

- Briefly mention the potential consequences faced by the teenager who threw the flower pot. This could be a court appearance, a community service requirement, or even counseling sessions, depending on the specific context. This adds closure and reinforces the notion of accountability for irresponsible actions.

3. Emphasize the social aspects of littering:

- Briefly discuss the negative impact of littering on the community as a whole. This could involve mentioning the financial burden of cleaning up litter, the potential for property damage, or the negative aesthetic impact on the neighborhood.

Here's an example of how the student could incorporate these suggestions:

"...Thankfully, the man behind the killer litter was successfully dealt with. Gopal and Mrs. Becca had never felt so relieved. Since that incident, the two of them would always walk along the sheltered walkways leading to their block. After all, it is better to be safe than sorry.

The incident left a profound impact on Gopal. Initially hesitant, his quick thinking and act of kindness averted a potential tragedy. Witnessing the consequences of littering firsthand, he developed a deeper understanding of the importance of individual responsibility. Mrs. Becca, initially shaken by the experience, felt a surge of gratitude towards Gopal and a renewed sense of safety within her community. The teenager, facing the consequences of his actions, hopefully embarked on a journey of self-reflection and learned a valuable lesson about responsibility and the impact of his choices.

Littering not only poses a physical threat to individuals but also carries a significant burden for the community. The cost of cleaning up discarded waste diverts resources from other crucial areas, while the presence of litter can damage property and create an uninviting environment for residents."

By incorporating these suggestions, the student can portray the human element and social impact of the incident more deeply, leading to a well-rounded and comprehensive exploration of the littering issue, potentially achieving a perfect score in content.

These steps were done for all the sample model essays (See Appendix B). Following that, all the generated feedback data was compared and a thematic analysis was conducted to pinpoint

similarities and differences across the feedback generated by Gemini for all the model essays of different themes (See Appendix C).

Findings and Discussion

An analysis of the data revealed recurring patterns across several aspects, though there were a few key differences.

Similarities and Differences Across Gemini's Output

One of the key similarities is that Gemini tends to give compositions the mark range of 16-18 for each section — this was seen for eight out of ten of the compositions. The other two compositions were given lower marks likely due to their lack of narrative structure or depth as seen in the feedback given. This was in itself quite interesting, given that the sampled essays were meant to be model essays and should not have such flaws.

There may be several reasons for Gemini's tendency to give the same range of marks. One reason is that, unlike automated writing systems that are trained with human-marked essays, Gemini does not have previous input to act as a model for marking thus it may not consistently have an idea of what a good essay might look like across different themes and topics. This might also reflect an intrinsic limitation of Gemini in evaluating nuanced writing. Additionally, Gemini accesses its information from a variety of sources like user contributions and publicly available web pages (Team Gemini, 2023), which might affect scoring due to biases that are unknown to the user (Slimi and Carballido, 2023). It could also be possible that Gemini is mimicking human marking behaviour – teachers are often quite conservative and reluctant to give a piece of writing perfect or near-perfect grades, in practice.

Looking at the feedback, there were patterns identified; most of which are in line with what Liu et. al. (2023) suggested that AI could generate feedback for — grammar, mechanics, and vocabulary usage. Table 1 below summarizes the main similarities across the feedback.

Table 1: Similarities across feedback from Gemini for both language and content areas

Language	Content
Show, don't tell	Deepen characters
Improve sentence structure	Deepen settings
Improve vocabulary and descriptions	Enhance ending/conclusion/resolution
Use literary devices (figurative language)	Explore the topic/theme more
Refine grammar (i.e., punctuation. phrases)	Avoid overgeneralization
Proofread/edit/revise	Improve essay structure/organization

We note that Gemini did not focus much on form across all the given feedback. This was expected, and also one of the reasons why we chose to sample model essays, which are generally grammatically sound, save for some instances of incorrect or missing punctuation and the wrong use of tense in some of the essays (Abe, 2008). If the feedback had focused on form and language errors instead, it might have led to a loss in clarity (on the students' part) in what good performance in essay-writing constituted – is it just to produce a grammatically sound piece of writing, or to develop a coherent and engaging story with clear structure. Gemini's feedback, instead, focused more on content and stylistic elements, despite claims in the literature that GAI was posited to struggle with issues like style and emotional depth (Oritsegbemi, 2023). In this sense, AI-generated feedback seems to adhere to Nicol & Macfarlane-Dick's (2006) principles of effective feedback, by clarifying what good performance is.

The feedback did at times focus on story structure, where specific feedback for each story section (introduction, body paragraphs, conclusion) was given, though the main focus seemed to be how to develop the content of the conclusion. Gemini's criteria for full marks for each section (content and language) were similar almost across all the compositions. This suggests that Gemini's idea and concept of an essay that is fully scored requires a lot more details than what was given within the various compositions, pointing to potential unrealistic expectations or biases in its evaluation.

A lot of the feedback centered on improving on the details: improve vocabulary and descriptions, use of literary devices, show don't tell, explore the topic/theme more, deepen characters and settings. Looking deeper into the specific feedback generated by Gemini, it was found that Gemini tended to give similar feedback with the same phrasing across different compositions. For example, in specific feedback for six out of ten of the essays, Gemini had suggested "show, don't tell". "Show, don't tell" is a common refrain in the Singapore writing classroom. It encourages students to actively engage with the text, builds their awareness of detail, and allows them to reflect on their work. A comparison can be seen in Table 2 below.

Table 2: Comparison of "Show, don't tell" feedback given by Gemini

	Composition title	Feedback header	Feedback (excerpts)
1	A Prank and Wake-Up Call	Show, don't tell	Instead of stating emotions directly (e.g., "Susan's ears grew red at the sight of such a humiliating thing."), use descriptive language to show the reader how Susan feels. For example, "Heat surged through Susan's face, matching the fiery red sign now hanging from her neck."
2	Traditional Festival that is Celebrated in Singapore	Show, don't tell	Instead of saying "I felt immense guilt," use descriptive actions and dialogue to show the reader the character's emotions.
3	Blackmail Blues	Show, don't tell	Help the student understand the concept of "showing" instead of "telling" their emotions and

			actions. This involves using vivid descriptions of characters, settings, and actions to allow the reader to infer the emotions and thoughts involved.
4	The Bully	Show, don't tell	Instead of directly stating emotions, use descriptive language and actions to show the characters' feelings. For example, instead of "I was terrified," describe the physical manifestations of fear like "My heart hammered against my ribs, and cold sweat prickled my skin."
5	The Jovial Janitor	Show, don't tell	Instead of directly stating emotions like "sadness" or "gratitude," use descriptive language that allows the reader to infer the characters' feelings through actions, expressions, and dialogue.
6	A Jovial Person	Show, don't tell	Instead of stating "Everyone loves being with him because of his positive outlook," show how his positive outlook impacts others through specific examples. Describe a situation where John's optimism helped someone overcome a challenge or uplifted their mood. This will make the writing more engaging and impactful.

From the feedback excerpts, it is clear that Gemini has a default or preferred feedback structure, namely, stating the change that should be made - "Show, don't tell" - followed by explaining its feedback using typically the same sentence stems - "Instead of...", "use descriptive language/action".

While the overall structure of Gemini's feedback might follow a similar format across different essays, its analysis goes beyond a one-size-fits-all approach. Gemini dives deeper, generating feedback that's specific to each piece. The ability for AI to provide personalized insight was explored by Liu et. al. (2023). Gemini can reference specific examples within the composition to support its points. This demonstrates a nuanced understanding of the individual student's work, rather than simply offering generic comments. This ability holds significant value for educators. Since Gemini personalizes its feedback to each student's work, it can act as a springboard for further analysis. Teachers can leverage this AI-generated output as a starting point. By utilizing the strengths and weaknesses identified by Gemini, educators can craft more specific and personalized responses for each student. This streamlines the initial assessment process, allowing teachers to focus on providing deeper and more impactful feedback. The ability to collate patterns in the feedback is also useful for teachers in a large classroom setting. As suggested by Fisher (2023), augmented marking takes place when educators use Gemini as a tool to expand on its analysis, enhancing the depth and quality of feedback provided to each student. This aligns AI-generated feedback with another one of Nicol & Macfarlane-Dick's (2006) principles of effective feedback, which is the ability to provide information to teachers that can be used for teaching.

One key difference found across all the feedback is also seen in the above example. Despite using the same prompt for all ten essays, there were differences in the way Gemini generated and structured feedback. There are instances where Gemini provides only indirect feedback. For example, in the case of the excerpt "The Bully," it might not always provide suggestions on how to "show, not tell" through specific sentence revisions. In such scenarios, educators might need to provide additional prompts to elicit more direct and granular feedback from Gemini. This could involve asking specific questions about areas like "show, don't tell" and prompting for illustrative examples. It's also important to acknowledge that Gemini's output can exhibit slight variations even with identical prompts. This variability stems from the inherent nature of large language models (Team Gemini, 2023). Educators should be aware of this and employ their expertise to evaluate and leverage the most relevant feedback generated by Gemini. By understanding these nuances, educators can effectively integrate Gemini's capabilities into their practice, maximizing its potential to enhance the feedback experience. A major implication for the implementation of AI in the classroom would thus be that teachers' AI literacy and the attendant training is indeed necessary for them to keep up with the dynamic changes brought by technology, as educators must understand how to follow up with specific prompts that will generate information they seek.

Another key similarity is the lack of personal pronouns and a lack of highlighting of strengths in the feedback generated by Gemini. It is posited that the personal pronouns "you" and "we" are used to personalize communication (Suryaningsih, 2021). While this may not be an important detail, Gemini's feedback does not sound personal and seems to increase social distance due to its lack of use of personal pronouns in feedback (Suryaningsih, 2021). Additionally, there is a lack of focus on the strengths of each composition. This is a classroom implication that teachers may want to note as positive teacher language is categorized as a dynamic element in nurturing learners' educational success (Sun, 2021). The use of positive teacher language during feedback may allow for the lowering of students' affective filter which in turn can allow for students to intake more comprehensible input (Krashen, 1982; Ni, 2012). The implication of this is that educators either have to prompt Gemini to reword its feedback in a positive tone or that they have to still take time to personalize and customize feedback to include positive teacher language to manage students' affective filters. In this respect, AI-generated feedback does not seem to fulfil one of Nicol & Macfarlane-Dick's (2006) principles of effective feedback, which is to encourage positive motivational beliefs and self-esteem.

Across all feedback output, Gemini also seems to focus on the structure of the compositions, namely the conclusions of the essays, with feedback that suggests that the resolutions should be expanded on. In Table 3 below, some examples of suggestions to strengthen the conclusion can be seen.

Table 3: Gemini's feedback on strengthening the conclusion

	Composition title	Feedback header	Feedback (excerpts)
1	Killer Litter	Expand on the resolution	Briefly mention the potential consequences faced by the teenager who threw the flower pot. This could be a court appearance, a community service requirement, or even counselling sessions, depending on the specific context. This adds closure

			and reinforces the notion of accountability for irresponsible actions.
2	A Prank and Wake-Up Call	Strengthen the conclusion	While the ending offers a good summary of the repercussions, consider adding a final thought or reflection to leave a lasting impression. Did this incident change the dynamic between Susan and Mr. Ravi? Did Susan develop better habits for completing her homework? Offer a glimpse into the potential lasting impact of this event.
3	A Naughty Prank	Strengthening the Ending's Impact	As suggested previously, the ending offers closure but lacks a deeper emotional connection. Consider revising the ending to create a lasting impression on the reader. You could show Sam reflecting on his actions and vowing to make amends, or even have him face a minor consequence beyond just paying Nicky.
4	Traditional Festival That is Celebrated in Singapore	Reflecting on the experience	Briefly mention how this experience impacted your understanding of the festival or your relationship with your family. Briefly mention how this experience impacted your understanding of the festival's traditions and their meaning. Did it deepen your appreciation for the cultural significance? Reflect on how this experience, despite the initial mishap, might have strengthened your relationship with your family. Did it teach you anything about communication, forgiveness, or family dynamics?
5	Chaotic Chinese New Year	Enhance the Resolution	Expand on the emotional impact: Show the characters' emotions instead of just stating them. Describe Aunt Susan's guilt through her trembling hands and tearful eyes. Show the family's concern for Grandma through their worried expressions and hushed voices. Develop the internal conflict: Explore the internal conflict within Aunt Susan. Did she genuinely believe comparing the children was best for them? Does she grapple with shame or regret for her actions?

			<p>Show, not tell: Instead of saying "settling of differences," showcase the reconciliation through actions and dialogue. Does Aunt Susan offer a sincere apology and hug her sister? Do the cousins share a playful moment, signifying a renewed bond?</p>
--	--	--	---

Gemini seems to put a lot of emphasis on the author's voice as much of the feedback is asking the author to include a reflection on what happened in the story (e.g. "Reflect on how this experience...", "Briefly mention how this experience impacted your understanding of") or for the author to include some sort of moral of the story that can be learnt from their composition. There is also a lot of emphasis on how to create a "lasting impact/impression" on readers. This finding is interesting as it seems that Gemini is prompting students to include more emotions in their stories — contrary to claims that AI typically does not focus on emotional depth (Oritsegbemi, 2023). This can be seen as evidence of AI-generated feedback delivering high quality information to students about their learning, another one of Nicol & Macfarlane-Dick's (2006) principles of effective feedback. They contend that where feedback is given, it is important that it is related to specific and relevant goals, standard or criteria – which seems to be the case here.

Following the rubric that it was fed, Gemini gave the compositions "Littering" and "A Jovial Person" lower scores compared to the rest of the compositions — 25 marks and 27 marks respectively. This suggests that Gemini can closely adhere to the marking rubric that was fed to it. This is in line with findings from Liu et. al. (2023) that GAI tools can analyze student work based on pre-defined parameters and provide personalized feedback.

The feedback generated by Gemini does aid in quickly identifying action points that can help students bridge the gap between their current and desired performance, which is another one of Nicol & Macfarlane-Dick's (2006) principles of effective feedback. There are specific actions such as "deepen character development", "adding sensory details" and "utilize figurative language" that Gemini generates along with guiding questions and examples. One example can be seen in the Killer Litter story where guiding questions are included in the feedback: "Can you describe the sounds of the city, the smell of the air after the flower pot fell, or the texture of Mrs. Becca's bag?" The use of questions is also encouraged in the STP (2024) to scaffold students' thinking instead of providing answers straight away. Sometimes, Gemini generates an example that students can use, as seen in feedback for Chaotic Chinese New Year, "Instead of just mentioning mess, describe the specific sights and sounds of the situation ("Scattered vegetables and fish lay in a colourful disarray, emitting a faint fishy aroma")."

However, there remains a few criteria of effective feedback in Nicol & Macfarlane-Dick's framework that AI-generated feedback does not naturally meet, such as encouraging teacher and peer dialogue around learning, and to a lesser extent, facilitating the development of self-assessment in learning. That the given feedback would foster further dialogue around learning between teachers and also peers is something that takes place beyond the GAI-assisted evaluation and feedback process, although it is possible that AI-generated feedback can feed that dialogic process. Additionally, although Gemini's feedback did not seem to overtly facilitate the development of self-assessment in learning, it is conceivable that specific prompt engineering beyond what we have tried in the current study can make it possible.

In closing, this study addressed two key research questions. The first is if Gemini could provide effective feedback on continuous writing tasks for primary school students in Singapore. This question was explored by analyzing Gemini's feedback on student essays. The findings

revealed that Gemini offers promise in several areas. It can deliver clear, age-appropriate feedback aligned with marking rubrics. It identifies weaknesses, suggests corrections, and proposes specific action points for improvement. This feedback helps educators pinpoint areas where students need to improve, ultimately informing their teaching strategies. However, limitations were also identified. While Gemini can generate examples of good writing, prompting might be necessary for specific instances. The sheer volume of suggestions might overwhelm students, requiring teachers to prioritize areas for improvement. Additionally, Gemini's feedback may lack focus on student strengths and personalization, and it might struggle with essays that deviate from traditional narratives.

The second research question addressed was the potential implications of GAI-generated feedback on the teaching and learning process in Singaporean primary school composition writing. Gemini offers valuable potential for educators who play a vital role in interpreting and tailoring Gemini's output to individual student needs. Gemini can streamline the initial assessment process by giving feedback that acts as a springboard for further feedback, allowing teachers to focus on providing deeper and more personalized feedback. While AI can identify strengths, weaknesses, and areas for improvement, teachers must tailor this feedback and address emotional factors that AI can't. Additionally, while Gemini can identify areas for improvement in narrative structure, its focus seems to be on conclusions. Thus, teachers may need to provide more guidance on the overall story structure, including elements like the inciting incident, climax, and pacing. To get the most out of AI feedback, educators will require training on using prompts effectively.

Limitations of Study and Future Directions

It is important to acknowledge that this study primarily focuses on methodology and potential applications. Further research is necessary to evaluate the effectiveness of this approach in actual classroom settings and to assess the impact on students' writing abilities. We agree with a reviewer's suggestion that future work could sample student-written essays, as well as get human (teachers) raters to evaluate the essays. Then, the AI-generated feedback can be compared against feedback provided by human raters. This comparison could offer valuable insights into the strengths and weaknesses of each approach, ultimately informing the development of more effective feedback strategies for primary school students' writing development.

Furthermore, future studies could explore refining the selection criteria for model essays and delve deeper into the types of AI feedback most beneficial for young learners. Additionally, investigating the integration of this method within existing curriculum frameworks and assessing teacher perspectives on its implementation would provide valuable insights for educational practice.

While the chosen five prompts offer a good starting point, future studies could consider expanding the scope to include additional writing genres and prompts. This could involve engineering prompts for specific curricular themes, character development, or creative writing exercises.

Conclusion

In conclusion, the integration of GAI such as Gemini, into the realm of writing feedback holds promise and challenges alike. As technology continues to redefine the boundaries of pedagogy, this research has delved into the potential implications for teachers and the effectiveness of

Gemini in providing feedback for primary school composition writing in Singapore. The intersection of traditional teaching methodologies with innovative AI solutions opens new avenues for educators to enhance and improve the writing instruction process. However, it is crucial to critically analyze the findings and recognize the limitations and considerations for practical implementation in the classroom.

One key aspect that emerges from this study is the time-consuming nature of the process, especially regarding the manual input of handwritten compositions into digital form. As students often produce hardcopy work, educators must invest significant time and effort in the transition to a softcopy format suitable for Gemini analysis. This practical limitation raises questions about the feasibility of implementing GAI in real classroom settings, where time is often a precious resource. Despite this challenge, the potential benefits of using Gemini in providing targeted and constructive feedback cannot be overlooked. The system's ability to assess language and content based on predefined rubrics, offer personalized insights, and identify areas for improvement demonstrates its potential as a valuable tool for educators.

Addressing the research questions posed in this study sheds light on the effectiveness of Gemini in providing feedback for primary school composition writing in Singapore. The first research question explored the extent to which Gemini can offer effective feedback on continuous writing tasks for primary school students. The findings reveal both promise and areas for improvement. Gemini demonstrates a tendency to provide consistent mark ranges that are mid to high, 16-18 marks, for different compositions, indicating a need for further refinement in understanding diverse themes and topics. The feedback generated often focuses on content and stylistic elements, emphasizing the importance of details, vocabulary enhancement, and narrative structure. Despite these patterns, Gemini showcases a general structure in providing feedback, allowing educators to leverage its insights as a basis for further personalized feedback.

In terms of content, Gemini consistently recommends improvements in areas such as deepening character development, enhancing sensory details, and utilizing figurative language. This aligns with the expectations of effective writing instruction, highlighting the importance of narrative elements in creating engaging and impactful compositions. Additionally, Gemini emphasizes the significance of the author's voice and suggests incorporating reflections or moral lessons to leave a lasting impression on readers. These insights into narrative and story structure reflect the alignment between Gemini's feedback and the underlying rubrics, indicating its ability to assess compositions based on the criteria fed into the system.

The second research question investigates the potential implications of GAI-generated feedback on the teaching and learning process in Singaporean primary school composition writing. Gemini's feedback has the potential to inform teaching practices by quickly identifying areas of improvement across multiple essays. The system aids in recognizing action points that can help bridge the gap between students' current and desired performance. This aligns with effective feedback principles, emphasizing timely and actionable insights to facilitate meaningful learning. Gemini's ability to generate feedback aligned with marking rubrics and offer valuable data for instructional adaptation positions it as a tool that can enhance teaching practices.

However, the study also reveals considerations for teachers when implementing GAI-generated feedback in the classroom. Gemini tends to provide feedback without personal pronouns and lacks a focus on highlighting strengths in compositions. The impersonal nature of the feedback and the absence of positive teacher language raises concerns about the potential impact on students' motivation and engagement. Educators need to be mindful of incorporating positive

feedback and personalization to create a supportive learning environment. While the system provides valuable insights, it should be seen as a complement to, rather than a replacement for, teacher expertise and guidance.

This research contributes valuable insights into the potential of GAI, specifically Gemini, as a tool for enhancing feedback delivery in Singaporean primary school composition writing. The findings suggest that Gemini can offer constructive feedback aligned with predefined rubrics, emphasizing narrative elements, details, and vocabulary enhancement. However, practical considerations, such as the time-consuming input process and the need for teacher intervention to personalize feedback, highlight the complexities of integrating GAI into the classroom. As technology continues to evolve, further research and refinement are essential to harness the benefits of GAI while addressing practical challenges, ultimately creating a symbiotic relationship between technology and traditional pedagogy in the realm of writing instruction. The journey toward effective AI-driven writing feedback in primary schools requires continuous exploration, collaboration, and adaptation to create a meaningful impact on student learning outcomes.

References

- Abe, Makoto. (2008). *Exploring the Role of Model Essays in the IELTS Writing Test: A Feedback Tool*.
- Bai, R., Hu, G., & Gu, P. Y. (2013). The relationship between use of writing strategies and English proficiency in Singapore Primary Schools. *The Asia-Pacific Education Researcher*, 23(3), 355–365. <https://doi.org/10.1007/s40299-013-0110-0>
- Bloom, B. S., Hastings, J. T., & Madaus, G. F. (1971). *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- Carless, D. & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment and Evaluation in Higher Education*, 43, 1315-1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Casco Publications. (2014). *How to Score English Oral & Model Compositions Primary 5*. Casco Publications Pte Ltd.
- ELIS. (2018). In *Assessment that impacts learning* (Vol. 6, pp. 1–13). Essay, English Language Institute of Singapore.
- Fischer, I. (2023). Evaluating the ethics of machines assessing humans. *Journal of Information Technology Teaching Cases*. <https://doi.org/10.1177/20438869231178844>
- Gemini Team, Anil, R., Borgeaud, S., Wu, Y., Alayrac, J. B., Yu, J., Bosma, M., Chen, C., Dehghani, M., Huang, P. S., Joulin, A., Lazzeri, F., Malmaud, J., Scialom, T., Schmid, C., Seidenschwarz, F., Wang, X., Zhai, X., ... Ahn, J. (2023). *Gemini: A family of highly capable multimodal models*. *arXiv preprint arXiv:2312.11805*. <https://doi.org/10.48550/arXiv.2312.11805>
- Gupta, R. (1995). Old beliefs impede student teacher learning of reading instruction. *Journal of Education for Teaching: International Research and Pedagogy*, 21(3), 347-360.
- Gupta, P., Ding, B., Guan, C., & Ding, D. (2024). Generative AI: A systematic review using topic modelling techniques. *Data and Information Management*, 100066. <https://doi.org/10.1016/j.dim.2024.100066>

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112. <https://doi.org/10.3102/003465430298487>
- Herold, B. (2022, April 12). Why schools need to talk about racial bias in AI-powered technologies. *Education Week*. <https://www.edweek.org/leadership/why-schools-need-totalk-about-racial-bias-in-ai-powered-technologies/2022/04>
- Ho, S. H., & Arkate, J. (2020). *English Thematic Model Compositions Primary 5*. Fairfield Book Publishers.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. New York: Prentice-Hall International.
- Lalande, J. F. (1982). Reducing composition errors: an experiment. *The Modern Language Journal*, 66(2), 140-149.
- Lee, I. (2022). *Creative Model Composition Primary 5*. Global Publishers.
- Lee, I. (2021). The Development of Feedback Literacy for Writing Teachers. *TESOL Quarterly*, 55, 1048-1059. <https://doi.org/10.1002/tesq.3012>
- Liu, M., Ren, Y., Nyagoga, L. M., Stonier, F., Wu, Z., & Yu, L. (2023). Future of education in the era of Generative Artificial Intelligence: Consensus among Chinese scholars on applications of chatgpt in schools. *Future in Educational Research*, 1(1), 72–101. <https://doi.org/10.1002/fer3.10>
- Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT toward lifelong learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4354422>
- Ministry of Education. (2013). Communication Slides for Assessment Literacy.
- Ministry of Education. (2023, September 20). *More support for schools and students to shape the future of learning*. Ministry of Education (MOE). <https://www.moe.gov.sg/news/press-releases/20230920-more-support-for-schools-and-students-to-shape-the-future-of-learning>
- Ni, H. (2012). The effects of affective factors in SLA and pedagogical implications. *Theory and Practice in Language Studies*, 2(7). <https://doi.org/10.4304/tpls.2.7.1508-1513>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Oritsegbemi, O. (2023). Human intelligence versus AI: Implications for emotional aspects of human communication. *Journal of Advanced Research in Social Sciences*, 6(2), 76–85. <https://doi.org/10.33422/jarss.v6i2.1005>
- Raffles Publications. (n.d.). *A Step by Step Guide to Ace Composition Writing Primary 5*. Raffles Publications.
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. John Wiley & Sons.
- Singapore Teaching Practice (STP). (2024, January 18). *Teaching Action: Comment Only Feedback*. Opal2.0.

- Slimi, Z., & Villarejo Carballido, B. (2023). Navigating the ethical challenges of Artificial Intelligence in higher education: An analysis of seven global AI ethics policies. *TEM Journal*, 590–602. <https://doi.org/10.18421/tem122-02>
- Sucipto. (2014). The Power of Feedback to Enhance Writing Skill. In *Annual Seminar on English Language and Teaching 2014* (pp. 423–429). Essay, English Department Faculty of Languages and Arts Universitas Negeri Padang. Retrieved from <https://ejournal.unp.ac.id/index.php/selt/article/viewFile/6736/5276>.
- Sun, Y. (2021). The effect of teacher caring behaviour and teacher praise on students' engagement in EFL Classrooms. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.746871>
- Suryaningsih, Y. (2021). A critical discourse analysis of personal pronouns in Greta Thunberg's speeches. *SAGA: Journal of English Language Teaching and Applied Linguistics*, 2(1), 55–64. <https://doi.org/10.21460/saga.2020.21.34>
- Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Systems with Applications*, 252A. <https://doi.org/10.1016/j.eswa.2024.124167>
- Wong, H. M., Kwek, D., & Tan, K. (2020). Changing assessments and the examination culture in Singapore: A review and Analysis of Singapore's assessment policies. *Asia Pacific Journal of Education*, 40(4), 433–457. <https://doi.org/10.1080/02188791.2020.1838886>
- Xu, Y., & Harfitt, G. (2018). Is assessment for learning feasible in large classes? challenges and coping strategies from three case studies. *Asia-Pacific Journal of Teacher Education*, 47(5), 472–486. <https://doi.org/10.1080/1359866x.2018.1555790>
- Yan, D., Zhang, S. (2024). L2 writer engagement with automated written corrective feedback provided by ChatGPT: A mixed-method multiple case study. *Humanities and Social Sciences Communication*, 11, 1086. <https://doi.org/10.1057/s41599-024-03543-y>
- Zhang, K., & Aslan, A. (2021). AI technologies for education: Recent research & future directions. *Computers & Education: Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/j.caeai.2021.100025>