rEFLections
Vol 31, No 3, September - December 2024

Generative AI and Essay Writing: Impacts of Automated Feedback on Revision Performance and Engagement

SUMIE CHAN
English Language Teaching Unit, The Chinese University of Hong Kong, China
NOBLE LO*
Faculty of Arts and Social Sciences, Lancaster University, United Kingdom
ALAN WONG
English Language Teaching Unit, The Chinese University of Hong Kong, China
Corresponding author email: p.k.lo@lancaster.ac.uk

Article information

Article history:
Received: 14 Jul 2024
Accepted: 12 Dec 2024
Available online: 18 Dec 2024

Keywords:
Automated feedback
Essay writing
Essay revisions
AI in education
Student engagement
Emotional responses

Abstract

This study investigates the impact of feedback generated by large language models (LLMs) on improving the essay-writing skills of first-year university students in Hong Kong. Specifically, it examines how generative AI supports students in revising their essays, enhances engagement with writing tasks, and influences their emotional responses during the revision process. The study followed a randomized controlled trial design, with one group of students receiving AI-generated feedback on their essay drafts while a control group did not. A mixed-methods approach was used to evaluate the feedback's effectiveness, combining statistical analysis of essay grades with student surveys and interviews. Quantitative results demonstrated that students who received AI feedback achieved significant improvements in essay quality, while qualitative findings revealed higher levels of engagement, increased motivation, and mixed emotional responses to the feedback process. These findings highlight the potential of generative AI as a tool for enhancing essay revision performance and fostering student engagement in higher education. However, further research is needed to explore its long-term impacts and applicability across diverse educational contexts.

INTRODUCTION

There is currently insufficient empirical evidence to conclusively support the claim that AI-generated feedback consistently improves students' revisions of academic work, particularly in the context of academic writing in Hong Kong higher education. However, several studies provide valuable insights into the potential and challenges of using AI in educational settings. For example, research suggests that while students in Hong Kong find AI tools convenient for learning writing skills, teachers express concerns regarding the generality and ambiguity of AI feedback, indicating the need for more precise and context-specific feedback mechanisms (Cheng, 2024). Studies on hybrid intelligence, which integrates AI and human input, reveal that formative feedback targeting individual errors can enhance students' writing skills, as

demonstrated in a study involving law students (Weber et al., 2024). Similarly, AI-generated corrective feedback has been shown to reduce writing anxiety and improve accuracy and fluency among language learners, outperforming traditional teacher feedback in some cases (Wang, 2024). Generative AI tools, although capable of enhancing academic work and providing learning feedback, require appropriate pedagogical support for effective implementation. Students are aware of both the benefits and risks of these tools, emphasizing the importance of careful integration into educational practices (Lee & Moore, 2024; Saúde et al., 2024). Moreover, while students generally perceive AI feedback as credible, they rate it lower in terms of goodwill compared to human feedback, which highlights the lack of a personal touch in AI-generated responses (Abendschein, 2024). Although AI-generated feedback shows potential in certain educational contexts, there is a lack of comprehensive empirical evidence supporting its consistent effectiveness in improving academic writing revisions in Hong Kong higher education.

Background

The field of Artificial Intelligence in Education (AIEd) has emerged as a rapidly growing area within educational technology, offering promising benefits for large-scale teaching contexts while delivering personalized, real-time feedback to learners (Gao et al., 2024). Although AI technologies have been integrated into education for over three decades, continued research is critical to advance intelligent support systems and address the challenges of large-scale implementation (Zawacki-Richter et al., 2019). Within this domain, Natural Language Processing (NLP), a branch of AI focused on the interaction between computers and human language, has made substantial progress, particularly with the development of transformer-based architectures. These architectures, which leverage self-attention mechanisms, have revolutionized text processing tasks. The emergence of large language models (LLMs), such as ChatGPT, points to a transformative potential for educational applications, enabling more sophisticated and context-aware interactions (Kasneci et al., 2023). As the technical capabilities of automated assessment systems have advanced, research has increasingly highlighted the potential of AIEd to enhance learning outcomes. Key developments in AIEd can be grouped into four primary areas (Zawacki-Richter et al., 2019): decision-making tools, intelligent tutoring systems, adaptive learning platforms, and systems for assessment and evaluation. These innovations collectively underscore the potential for AI-driven tools to fundamentally reshape the education landscape.

Decision-making tools play a critical role in various educational processes, including profiling and predicting outcomes such as admissions decisions, course scheduling, student retention and drop-out rates, academic performance, and student modeling (Alvero et al., 2020; Chen et al., 2020; Langley, 2019). Intelligent tutoring systems, on the other hand, are designed to deliver course content, engage in interactive learning with students, curate tailored learning materials, promote collaboration, and provide support to educators (Feng & Law, 2021; Hwang et al., 2020). Adaptive systems focus on offering scaffolding and personalizing content to meet individual learning needs. These systems assist teachers in understanding student progress, utilize academic data to monitor and guide learners, and visually organize knowledge through tools such as concept maps (Chen & Bai, 2010; Kabudi et al., 2021). Similarly, assessment and evaluation tools have advanced significantly, enabling automated grading, delivering feedback,

evaluating student comprehension and engagement, promoting academic integrity, and assessing the effectiveness of teaching practices (Huang et al., 2023; Luckin, 2017). Together, these four areas—decision-making tools, intelligent tutoring systems, adaptive systems, and assessment tools—represent substantial progress in the field of AIEd and highlight its potential to revolutionize educational practices.

This study is grounded in the potential of feedback to enhance student performance in written assignments. Numerous studies have highlighted the significant role feedback plays in fostering student learning and academic improvement (Graham et al., 2015). Additionally, research has shown that feedback specifically aimed at guiding the revision process can lead to better outcomes, including higher grades on written assignments (Gnepp et al., 2020). However, providing detailed and effective feedback is often a labor-intensive task for educators, with grading frequently cited as a major contributor to teacher workload and stress (Hahn et al., 2021). At the same time, university students commonly report dissatisfaction with the quality of feedback they receive (Madigan & Kim, 2021). These challenges underscore the need for strategies that can both enhance the quality of feedback and alleviate the burden on teachers. Automated feedback systems offer a promising solution to achieving both goals simultaneously (Gao et al., 2024).

Automated feedback systems hold significant potential for improving feedback consistency while simultaneously alleviating teachers' workload. Several studies have already explored the use of automated writing evaluation (AWE) tools to reduce the grading burden on educators (Crossley et al., 2022). Moreover, there is an expanding body of research examining how automated feedback can be effectively implemented through computer programs (Fleckenstein et al., 2023). However, many earlier efforts to develop such systems have primarily focused on task-specific applications, which inherently limit their adaptability. These systems often struggle to accommodate courses where teachers assign a variety of tasks, such as offering students multiple essay prompts to choose from. Additionally, they face challenges in contexts where assessments rely on broad or subjective criteria, such as opinion-based or reflective writing tasks (Ramesh & Sanampudi, 2022).

Rationale

The rapid advancements in AI technologies during the early 2020s have opened up new possibilities for leveraging generative AI powered by large language models (LLMs) to both evaluate written assignments and deliver feedback aimed at enhancing students' writing skills. As highlighted in the following literature review, research into the use of generative AI for providing feedback on student work is steadily expanding. However, the varied nature of assessment practices across different educational systems and cultural contexts often raises concerns about the generalisability and transferability of findings. In particular, there remains a lack of sufficient empirical evidence to substantiate claims that AI-generated feedback consistently improves students' revisions of academic work, especially within the context of academic writing in Hong Kong. By conducting research in this area, this study seeks to evaluate the practical utility of generative AI based on LLMs and to assess its relevance and effectiveness in contemporary university-level academic writing in Hong Kong.

Aims

This study seeks to address the research gap concerning the use of AI in providing feedback on essay writing in the context of Hong Kong. Its objectives are twofold: (1) to evaluate the potential of AI in delivering feedback on the essay revisions of university-level language students in this specific setting, and (2) to contribute to the growing body of literature on how AI technologies are transforming academic writing instruction and assessment practices. By achieving these aims, the study intends to make a meaningful contribution to the development of effective AI-driven feedback mechanisms for essay writing within Hong Kong's universities, while also advancing the broader conversation on the empirical foundations for using AI to support academic writing across diverse educational contexts.

Research questions

Achieving these objectives requires the development of a study specifically designed to address the identified knowledge gap. To this end, the research focuses on the following questions:

1. To what extent can LLM-based generative AI provide feedback on written assignments that enhances the quality of students' work?
2. What are students' experiences when receiving feedback from LLM-based generative AI, particularly regarding their motivation, emotional responses, and attitudes?

To answer these questions, the study adopts a mixed-methods approach, combining both quantitative and qualitative data collection and analysis. The methodology section below provides a detailed explanation of the experimental design, along with the rationale for the chosen methods and research instruments.

LITERATURE REVIEW

The existing literature on the use of LLM-based generative AI for providing feedback highlights its potential as a valuable tool in education. Earlier studies primarily concentrated on automated writing evaluation (AWE) systems and their ability to assess student work. However, these studies often pointed out the limited capacity of such systems to deliver personalized or tailored feedback (Mertens et al., 2022). More recent research into the effectiveness of LLM-based generative AI in enhancing student performance on revised assignments has produced mixed results. Challenges have been noted regarding the applicability and specificity of the feedback, as it often remains constrained to the types of tasks or criteria the AWE systems were explicitly designed to evaluate (Fleckenstein et al., 2023).

In contrast, large language models (LLMs) such as GPT have recently emerged as tools capable of offering more customized feedback on written assignments (Yang et al., 2023). Trained on extensive datasets of textual information, LLMs are designed to produce natural language responses that closely resemble human-generated feedback (Bowman, 2023). Unlike traditional automated writing evaluation (AWE) systems, LLMs can adapt their feedback to

various types of tasks by considering specific inputs, learning objectives, and scoring criteria, all while requiring significantly less coding and development time (Bressane et al., 2024). This adaptability highlights the considerable potential of LLMs, such as GPT, to revolutionize the delivery of automated feedback in educational contexts (Wardat et al., 2023).

Despite their potential, there is still a notable lack of empirical evidence supporting the effectiveness of feedback generated by LLMs, as highlighted by Tao et al. (2024). Concerns have been raised regarding the accuracy of AI-generated feedback, particularly given that generative AI models like GPT are prone to making factual errors when performing certain tasks (Lee et al., 2024a). Additionally, researchers have pointed out that the effectiveness of AI feedback may diminish outside of controlled research settings. For instance, when used by non-experts, the quality of feedback can be compromised if students fail to provide sufficiently detailed or clear prompts (Knoth et al., 2024). Nevertheless, LLMs tend to excel in creative tasks, which may include generating nuanced and adaptive feedback (Chia et al., 2023). Furthermore, Tao et al. (2024) emphasize that LLMs can provide feedback without relying on reference texts, demonstrating greater versatility and potential for feedback generation compared to traditional AWE systems.

The potential of LLMs to deliver effective feedback is reinforced by findings from several empirical studies. For instance, one study comparing feedback generated by LLMs with that provided by instructors on university students' written reports found that the AI-generated feedback was coherent and largely aligned with instructor feedback in terms of positive or negative evaluations of the work (Dai et al., 2023). Similarly, studies that rely on evaluations from students or instructors have reported favorable perceptions of LLM-generated feedback, with participants acknowledging its utility and effectiveness (Jacobsen & Weber, 2023; Steiss et al., 2024). However, these positive assessments are tempered by concerns raised in research on English language teachers' perspectives. Some educators have expressed reservations about issues such as linguistic accuracy, the potential for students to become overly dependent on AI, and the risk of stifling student creativity (Al-Khreseh, 2024). Despite these insights, there remains a notable lack of empirical studies specifically measuring the impact of generative AI on student learning outcomes.

The limited studies conducted so far suggest encouraging outcomes regarding the potential of generative AI to provide valuable feedback on student work. For instance, one study investigating GPT-generated feedback found that students who incorporated AI into their research and revision processes, including using it for feedback, outperformed their peers in critical, reflective, and creative thinking skills compared to those relying on traditional methods of research and feedback (Essel et al., 2024). Similarly, Meyer et al. (2024) conducted a study involving 459 upper secondary EFL students, dividing them into two groups: one group received feedback generated by an LLM, while the other did not. The results indicated that the group receiving AI-generated feedback and revising their work showed greater improvement in the quality of their written assignments compared to the group relying on non-AI feedback. This finding highlights the potential applicability of such tools in enhancing ESL student outcomes at the university level, suggesting a promising avenue for further exploration in higher education contexts.

Beyond the immediate learning outcomes, such as improvements in the scoring of written work, there are additional dimensions where AI-generated feedback can be compared to instructor feedback. For example, research indicates that students' perceptions of the value of completing certain English as a Second Language (ESL) writing tasks are strongly linked to their motivation to engage with these tasks (Eccles & Wigfield, 2020). Motivation, in turn, is closely tied to positive emotional states, which play a critical role in the writing process (Schrader & Kalyuga, 2020). Instructor feedback has been shown to foster these positive emotions, which can enhance students' motivation and overall engagement with writing tasks (Lipnevich et al., 2021). However, for feedback—whether from an instructor or AI—to effectively influence emotions, motivation, and task engagement, students must perceive it as meaningful and effective (Pandero & Lipnevich, 2022). This highlights the importance of understanding how students evaluate feedback from different sources and its broader impact on their learning experience.

Preliminary evidence suggests that LLM-based feedback can positively influence student emotions. For example, a study by Li and Xing (2021) found that LLMs were capable of providing effective emotional support to students, helping to foster a more positive learning experience. Similarly, research by Aslan et al. (2024) demonstrated that interactions with generative AI elicited favorable perceptions and high levels of engagement, though this particular study was conducted with younger learners. In another study, Zheng and Stewart (2024) examined the use of GPT with EFL students and noted that the model's ability to reframe tasks contributed to fostering greater cultural awareness among participants, who responded positively to the approach. Additionally, Al Shloul et al. (2024) explored the use of GPT in providing feedback to improve student performance. Their findings revealed that the majority of students valued the feedback provided by the AI and found the interaction both engaging and beneficial.

What remains less understood, despite the findings of the aforementioned studies, is whether LLMs can deliver feedback in a manner that students consistently perceive as effective, while simultaneously fostering positive emotions and motivating them to engage with their work. This question is particularly significant given that the feedback-revision-submission cycle can often be emotionally taxing and demotivating for some students (McGarrell & Verbeem, 2007). While Meyer et al. (2024) observed moderate increases in task motivation and positive emotions among participants receiving LLM-generated feedback, this suggests only part of the potential emotional impact of such systems. To build on these findings, there is a growing need for qualitative research to explore these relationships in greater depth. Such research could provide valuable insights into the specific aspects of AI-generated feedback that students perceive as helpful, motivating, or emotionally supportive, as well as those that may elicit negative reactions.

This literature review has identified several key gaps in the existing body of research. First, while a limited number of studies have explored the effectiveness of LLM-based generative AI on student outcomes, the findings, though promising, remain sparse. These studies generally report positive correlations between the use of AI-generated feedback and improvements in student learning outcomes following the revision of work. However, there is a notable lack of

research examining whether AI-generated feedback can positively influence student revisions in the specific context of education in Hong Kong. This underscores the need for further studies that address how generative AI can be applied within specific educational environments and areas of learning.

Additionally, while early evidence suggests that LLM-generated feedback may help foster positive emotions and motivation among students, more qualitative research is needed to uncover the underlying mechanisms driving these effects. Understanding how students engage with and respond emotionally to AI-generated feedback will enable a deeper comprehension of its potential role in improving learning experiences. These identified gaps have directly informed the design of this study, which seeks to address these issues and is outlined in the following sections.

METHODOLOGY

Sample

The study involved data collection from 918 first-year students enrolled in an English-language course at a higher education institution in Hong Kong. All participants were taking a course that incorporated the International English Language Testing System (IELTS) as part of a foundational university writing skills program. The sample exclusively included students who were Hong Kong citizens with English as their second language, while students who did not meet these criteria were excluded from the study.

Of the sample, 55% were female and 45% were male, and efforts were made to ensure that the control and experimental groups reflected this gender distribution as closely as possible. The sample was divided into two groups: 342 students were assigned to the feedback group, which received AI-generated feedback, while 576 students formed the control group, which did not receive AI-based assistance. This division allowed for a comparative analysis of the impact of AI-generated feedback on student outcomes and experiences.

Experimental design

The experiment was conducted during a two-hour lesson held in a computer laboratory on the university campus. Participating students were required to complete the following writing task under test conditions:

> *Do you agree or disagree with the following statement? Children under five ought to be prohibited from using tablet computers or smartphones. Use specific reasons and examples to support your answer.*

To maintain academic integrity, a researcher was present throughout the session to monitor students, ensure adherence to test conditions, and prevent plagiarism. Additionally, the researcher facilitated the use of generative AI tools as part of the experimental design.

Students were allocated 30 minutes to complete the writing task. After finishing, they were instructed to email their written responses directly to the researcher for further analysis. This structured environment ensured consistency in task conditions across all participants.

Those in the feedback group had their work submitted by the researcher to GPT 3.5. This was preceded by a prompt setting out the task instructions and learning objectives and requesting no more than 500 words of feedback (app. 1):

A number of undergraduate students studying English at a university in Hong Kong have been tasked to write an essay on the following question:

> 'Do you agree or disagree with the following statement? Children under five ought to be prohibited from using tablet computers or smartphones. Use specific reasons and examples to support your answer.'

They have been tasked with meeting the following four learning objectives upon completion of the assignment:

> '1. Engage critically with the question, meeting the criteria set out in the instructions.
> 2. Argue convincingly towards a clear thesis, drawing on existing research or evidence where possible.
> 3. Employ academic vocabulary and formal English in your answer.
> 4. Ensure accurate spelling and grammatical coherence throughout.'

Based on these learning objectives, please provide no more than 500 words of feedback designed to help the student improve their essay and raise their grade, giving both general and specific guidance as to areas and means for potential improvement. The essay in question follows below.

All participating students received an email instructing them to revise and improve their essays. Students in the feedback group were provided with the LLM-generated feedback from GPT-3.5, while those in the control group received generalised guidance that had been pre-prepared by a teacher. This generalised feedback outlined broad strategies for improving academic writing but did not include specific or tailored comments.

Students were given 5 minutes to prepare and review their feedback before beginning the revision process. They were then allotted 20 minutes to revise their work, after which they resubmitted their revised essays to the researcher for evaluation.

Both the original and revised essays were assessed by instructors from the course. Manual scoring was employed due to concerns regarding the limited accuracy of LLMs, including GPT, in reliably grading student writing (Lee et al., 2024b; Misiejuk et al., 2024). To ensure fairness and consistency, each essay and its corresponding revision were marked by separate instructors. Furthermore, all essays were double-marked, with the final score calculated as the average of the two independent assessments. This rigorous marking process ensured reliability and minimized potential bias in the evaluation of student performance.

Quantitative methods

At the conclusion of the experiment, all participating students were asked to complete a brief questionnaire about their experiences during the revision process. The questionnaire primarily focused on their emotions and engagement with the task, asking students to rate how positive their emotions were while revising, how motivated they felt to complete the revisions, and how engaged they were with the overall revision process (see Appendix 2). Scalar responses were collected, enabling comparisons between the feedback and control groups, as well as the relationship between these responses and their scores from the writing task.

The data collected from the questionnaires were analyzed using IBM's Statistical Package for the Social Sciences (SPSS) 29.0. This software facilitated the definition of variables (e.g., gender, scalar variables, etc.) and the creation of cases based on data entered into the program (Salcedo & McCormick, 2020). To examine relationships and differences within the data, statistical tests were conducted. For instance:

- Pearson's product-moment correlation coefficient was used to identify numerical relationships between sets of data, such as the correlation between emotional positivity and test performance.

- The Student's t-test was applied to compare the means of two or more groups' scores across numerical variables, such as writing task scores or scalar responses from the questionnaire (McCormick & Salcedo, 2015).

An alpha level of 0.05 was used as the threshold for statistical significance across all tests. These analyses were applied to variables such as test scores, questionnaire responses, and other relevant data to determine whether significant differences or relationships existed between the groups.

Qualitative methods

Following each experiment session, an interview was conducted with a participating student from the feedback group to discuss their experience with LLM-generated feedback. In total, 16 interviews were successfully completed, each lasting approximately one hour. Interviews were chosen as the qualitative method due to their ability to generate in-depth insights into individual perspectives, which cannot always be captured effectively through questionnaires (Peters & Halcomb, 2015).

The interviews were conducted by the researcher using a semi-structured approach. This method allowed the researcher to follow a set of predetermined questions while also providing the flexibility to probe deeper into areas of interest or clarify responses, ensuring a richer understanding of student experiences (Magaldi & Berler, 2020). Each interview was recorded using digital audio recording software on the researcher's tablet computer. The recordings were then transcribed automatically using digital transcription software and subsequently reviewed and manually corrected to address any errors in the transcription process.

The transcribed interview data were subjected to thematic analysis, a method used to identify, analyze, and report patterns or themes within qualitative data (Attride-Stirling, 2001). Thematic analysis is particularly useful for describing and interpreting recurring themes within a dataset, offering a structured approach to understanding interviewees' perspectives (Braun & Clarke, 2006).

To facilitate the coding and analysis of the data, the software tool Leximancer was employed. Leximancer uses machine-learning algorithms to extract semantic and relational data from the text, automatically grouping these into themes (Smith & Humphreys, 2006). The software generates outputs such as heat maps that visually represent the prominence and relationships between themes, as well as ranked and co-occurring concepts, providing a comprehensive overview of the data (Smith & Humphreys, 2006).

An unsupervised approach to coding and data analysis was adopted, allowing the researcher to take an inductive approach. By not imposing pre-existing categories or frameworks, this approach ensured that the analysis was guided by the data itself, enabling the discovery of emergent themes and insights based on the students' responses to the interview questions.

Ethical considerations

Ethical considerations were carefully addressed in the design and implementation of this research. The British Educational Research Association's Ethical Guidelines for Educational Research (British Educational Research Association, 2018) were consulted to ensure adherence to established ethical standards. In line with these guidelines, all participants took part in the study on a voluntary basis and were fully informed of their rights, including the option to withdraw from the study at any time without penalty.

To protect the privacy and confidentiality of participants, their data was anonymized at the point of marking and transcription. Each participant's work was assigned a codename (e.g., Student 1, Student 2, etc.), ensuring that no personally identifiable information was associated with their submissions or responses.

Furthermore, the researcher was mindful of the power dynamics inherent in their relationship with the students, as this can influence participants' behavior and responses in educational research. Reflecting on the importance of positionality in qualitative research, the researcher took steps to minimize any potential bias or undue influence throughout the study (Holmes, 2020). This included fostering an environment of trust and neutrality during interactions with participants.

By addressing these ethical considerations, the study sought to maintain the integrity of the research process while protecting the rights and well-being of all participants.

RESULTS

Quantitative analysis

The results of the tests and questionnaires reveal notable differences between the feedback and control groups. As shown in Table 1, the revised scores for the feedback group were higher than those for the control group. Specifically, the feedback group achieved an average of 3.113 additional marks on their revised papers compared to the control group.

In addition to improvements in test scores, the feedback group reported higher levels of positive emotions, motivation, and engagement during the revision process. Among these variables, motivation stood out as particularly significant, with the feedback group reporting an average increase of 1.7 points on a scalar scale of 1 to 10 compared to the control group. These findings suggest that LLM-generated feedback not only enhanced students' academic performance but also positively influenced their emotional and motivational responses to the revision process.

Table 1

Average scores for task and questionnaires across and between feedback and control groups

Group	Task Score	Rev. Score	Difference	Emotion	Motivation	Engagement
Feedback	56.40206	63.98969	7.587629	4.556701	5.051546	4.969072
Control	56.68041	61.15464	4.474227	3.762887	3.350515	3.989691
Difference	-0.27835	2.835052	3.113402	0.793814	1.701031	0.979381

Applying the t-test to the test scores and self-reported questionnaire results allowed for an assessment of whether there were statistically significant differences between the feedback and control groups (see Table 2).

In terms of improvements in test scores, the feedback group demonstrated a mean improvement of 7.588 marks (SD = 7.477), compared to a mean improvement of 4.474 marks (SD = 7.157) in the control group. The resulting p-value of 0.003604 indicates a statistically significant difference, suggesting a high likelihood that the null hypothesis can be rejected. This supports the conclusion that LLM-generated feedback had a meaningful impact on improving student performance during the revision process.

For self-reported positive emotion scores, the feedback group had a mean score of 4.557 (SD = 3.07), compared to a lower mean score of 3.763 (SD = 2.761) for the control group. The difference of 0.794 points was the smallest among all self-reported metrics and was not found to be statistically significant (p = 0.0611).

In contrast, the differences in motivation and engagement scores were more pronounced. The feedback group reported a 1.701-point higher mean motivation score (feedback group SD = 3.108; control group SD = 2.479), which was found to be statistically significant (p = 0.00004). Similarly, the feedback group demonstrated a 0.979-point higher engagement score (feedback group SD = 3.411; control group SD = 2.951), with this difference also proving statistically significant (p = 0.0346).

These results suggest that while LLM-generated feedback did not significantly influence students' positive emotions during the revision process, it had a statistically significant impact on test performance, as well as on their motivation and engagement levels. These findings highlight the potential for LLMs to enhance both academic outcomes and students' experiences of the writing and revision process.

Table 2

T-test results

	mean diff.	df	t-value	p-value
Improvements	3.113	192	2.947	0.0036
Emotion	0.794	192	1.884	0.0611
Motivation	1.701	192	4.193	0.00004
Engagement	0.979	192	2.089	0.0346

The size of the effects between the feedback and control groups can be further examined by calculating coefficient scores, which provide insight into the strength of the relationships between receiving feedback and the observed outcomes. To explore the relationship between receiving AI-generated feedback and the improvement in scores between the original and revised papers, a point-biserial correlation was calculated. This analysis produced a coefficient of 0.208, indicating a weak positive correlation between receiving AI-generated feedback and greater improvements in test scores.

When comparing the feedback and control groups across self-reported measures, the correlations were weaker. A very weak positive correlation was observed between receiving feedback and positive emotion (0.135) and also between receiving feedback and engagement (0.152). However, the correlation between receiving feedback and motivation was larger, with a weak-to-moderate positive correlation of 0.29.

Overall, the results suggest that receiving AI-generated feedback is positively correlated with improvements in test scores and all self-reported measures (positive emotion, motivation, and engagement). However, the strength of these correlations varied, with the strongest relationship observed for motivation. It is important to note that while positive correlations were found, the statistical significance of these relationships must be considered to fully interpret their implications.

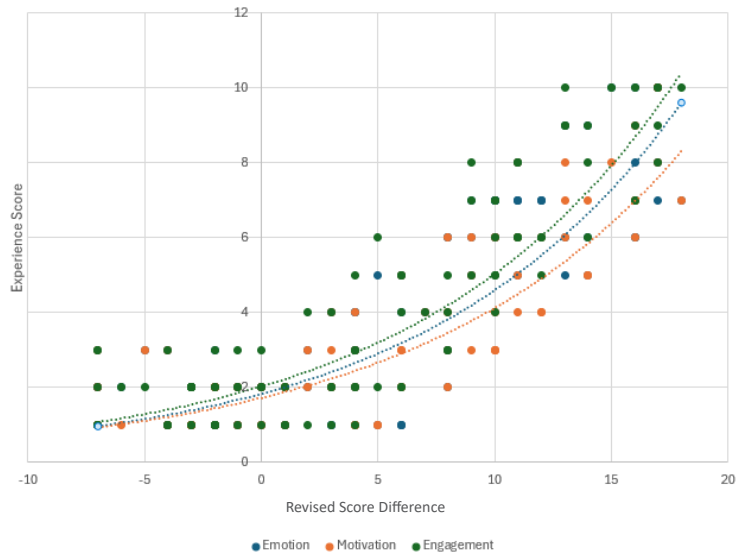


Figure 1 Self-reported scores for experiences (positive emotions, motivation, and engagement) cross-referenced with associated score improvement on revised task submissions

Attempting to identify the mechanisms underlying the observed differences between the feedback and control groups requires examining how variables such as emotion, motivation, and engagement correlate with improvements in test scores. To explore these relationships, Pearson's correlation coefficient was calculated to measure the strength and significance of the effects of these variables on the difference between original and revised paper scores.

When analyzing the group as a whole, the correlation between emotional scores and the improvement in test scores was $r = 0.543$, indicating a moderate positive effect that was statistically significant ($p = 1.83188 \times 10^{-15}$). The effects of motivation and engagement were considerably stronger, with $r = 0.882$ and $r = 0.883$, respectively. These represent very strong positive correlations, both of which were highly statistically significant ($p = 1.24235 \times 10^{-64}$ and $p = 5.90325 \times 10^{-65}$, respectively).

As illustrated in Figure 1, the relationships between the three affective variables (emotion, motivation, and engagement) and improvements in test scores appeared to increase exponentially, suggesting that as students experienced more positive affective states during feedback and revision, their performance gains became increasingly pronounced. This finding highlights the critical role of positive emotional and motivational states in driving successful revision outcomes.

It can therefore be theorized that the strong relationships between motivation and engagement and the improvements in test scores, when paired with the weaker effect of receiving feedback on motivation and engagement, may explain the larger relative improvement in the revised paper scores observed among the feedback group. In other words, while AI-generated feedback had only a modest direct influence on motivation and engagement, these factors appear to act as mediators that amplify the impact of feedback on revision outcomes.

To confirm these findings, additional tests were conducted to measure the effects of engagement and motivation on test score improvements. Simple linear regression analyses were performed separately for both the feedback and control groups (see Tables 3–6).

For the effect of motivation on score improvements, a strong correlation was observed in the control group, with a Multiple R of 0.842 and an R-Square value of 70.1%, indicating that motivation accounted for a significant proportion of the variance in score improvements. However, the correlation was even stronger for the feedback group, where motivation accounted for an additional 10% of the variance in score differentials compared to the control group. This suggests that the feedback group benefited more from the motivational effects of the revision process.

When examining the effect of engagement on score improvements, similar results were found. For the control group, the Multiple R was 0.873 with an R-Square value of 76.2%, indicating a strong correlation between engagement and score improvement. In the feedback group, the Multiple R increased slightly to 0.889, with the R-Square rising to 79%, demonstrating a slightly stronger relationship between engagement and score improvements.

Interestingly, the feedback group experienced a larger increase in the effect of motivation than engagement, as reflected by the more pronounced improvement in R-Square values for motivation. These findings suggest that AI-generated feedback has a greater influence on motivation compared to engagement, amplifying the motivational benefits of the revision process to a greater extent.

Table 3

Regression analysis of motivation effect on score difference for control group

Multiple R	0.842
R-Squared	70.1%
Standard Error	1.351
Significance F	3.388×10^{-27}

Table 4

Regression analysis of engagement effect on score difference for control group

Multiple R	0.873
R-Squared	76.2%
Standard Error	1.407
Significance F	2.07×10^{-31}

Table 5

Regression analysis of motivation effect on score difference for feedback group

Multiple R	0.911
R-Squared	83%
Standard Error	1.293
Significance F	2.19×10^{-38}

Table 6

Regression analysis of engagement effect on score difference for feedback group

Multiple R	0.889
R-Squared	79%
Standard Error	1.581
Significance F	6.38×10^{-34}

Qualitative analysis

Thematic analysis of the interviews with sixteen students was conducted using Leximancer, a software tool that employs an in-built algorithm to code and organize themes. Leximancer identifies themes based on the frequency, proximity, and semantic connections between terms used in the interview transcripts. This approach allows for a data-driven, objective analysis of the key ideas and patterns present in the qualitative data.

A concept map (Figure 2) was generated to visually represent the relationships among the concepts identified in the interviews, as well as their associations with related concepts. Six main concepts emerged from the analysis: feedback, paper, having, task, better, and forward. These concepts reflect the recurring ideas and themes discussed by participants during the interviews.

The prevalence of these concepts is further detailed in Table 7, which provides a quantitative breakdown of their frequency across the interviews. This table clarifies the relative importance of each concept and offers insights into the dominant themes in students' reflections on their experiences with LLM-generated feedback.

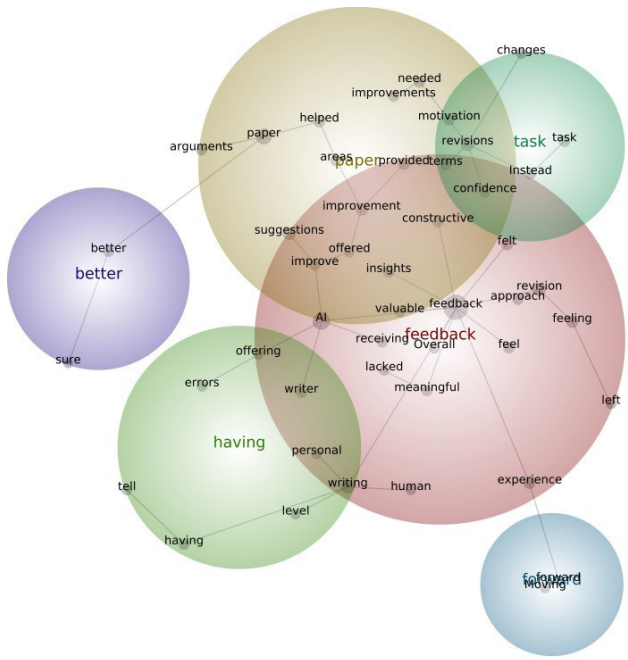
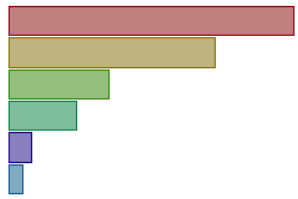


Figure 2 Concept map for analysis of interviews derived from *Leximancer* analysis

Table 7

Table of major themes with frequency of conceptual frequency

Theme	Hits
feedback	208
paper	150
having	73
task	49
better	16
forward	10



These thematic concepts, as identified in Table 7, also encompass sub-concepts, the frequencies of which are detailed in Table 8. Some related concepts were grouped together during analysis—providing further insight into the relationships between ideas expressed by participants. For example, sub-concepts such as 'writing', 'revision', 'felt', 'feel', and 'experience' were grouped under the broader theme of 'feedback'. Similarly, sub-concepts such as 'suggestions', 'help', 'motivation', 'provided', and 'improved' were categorized under the theme of 'paper'.

These groupings illustrate the semantic relationships between concepts in the students' responses, highlighting how specific ideas are linked to broader themes. The connections between these concepts suggest that participants frequently associated certain experiences or actions (e.g., receiving suggestions or feeling motivated) with specific themes, such as feedback or the process of improving their papers.

This analysis is further developed below by examining concept pathways—the patterns of relationships between concepts—and the excerpts attached to certain themes identified during the analysis. These excerpts provide qualitative evidence to support the patterns revealed in the concept maps and frequency tables, offering deeper insight into the ways students experienced and processed the feedback they received.

Table 8

Ranked concepts derived from Leximancer analysis of interviews with participating students

	Count	Relevance				
feedback	193	100%	provided	17	9%	
paper	77	40%	insights	17	9%	
writing	55	28%	improve	17	9%	
revisions	40	21%	constructive	17	9%	
revision	36	19%	errors	16	8%	
suggestions	29	15%	left	14	7%	
felt	26	13%	tell	14	7%	
helped	25	13%	offered	13	7%	
feeling	25	13%	meaningful	13	7%	
improvement	23	12%	improvements	13	7%	
task	23	12%	confidence	13	7%	
feel	23	12%	arguments	13	7%	

	Count	Relevance				
experience	23	12%		terms	12	6%
having	23	12%		approach	12	6%
areas	20	10%		personal	12	6%
needed	20	10%		writer	12	6%
motivation	19	10%		valuable	11	6%

With respect to the theme of 'feedback', respondents used a range of related concepts, including AI, writing, revision, feeling, felt, insights, feel, experience, meaningful, valuable, receiving, approach, left, human, and lacked. These terms highlight the emotional connections participants made with the experience of receiving AI-generated feedback. While many emotional responses were positive, some were notably negative.

In some cases, participants expressed dissatisfaction with AI feedback, describing it as impersonal and disengaging. They noted that AI lacked the empathetic and personalized touch they associated with feedback from human instructors or peers. For them, this absence of human interaction diminished the motivational impact of the feedback.

On the other hand, several participants found AI feedback helpful and encouraging. For instance, they felt the feedback provided them with a sense of direction, helping them understand how to improve their work or what steps to take next. Some participants remarked that the feedback motivated them to strive for better results and take their revisions more seriously.

However, overall emotional responses to AI feedback were mixed. While some appreciated the guidance it provided, others expressed a clear preference for human feedback, which they felt was more supportive and personal. Only one participant indicated that they found AI-generated feedback superior to human feedback, reflecting the general skepticism among participants about the emotional value of AI as compared to human interaction.

The connections between emotional concepts and other concepts shed light on how participants' emotional responses to AI feedback sometimes related to improved outcomes. As shown in Figure 3, terms such as 'feeling' and 'improvement' were frequently linked through concepts like 'suggestions' and 'helped', which were among the most commonly mentioned ideas. For instance, participants often reported feeling more capable of improving their submissions or clearer about the revisions they needed to make after receiving feedback.

That said, these connections between emotional responses and improved outcomes were generally weaker than the connections between emotions and specific emotional experiences. Participants' immediate emotional reactions, as captured by concepts such as 'sad', 'delighted', and 'irritated', were more prominent in the data, reflecting the varied and often conflicting feelings participants had about using AI-generated feedback.

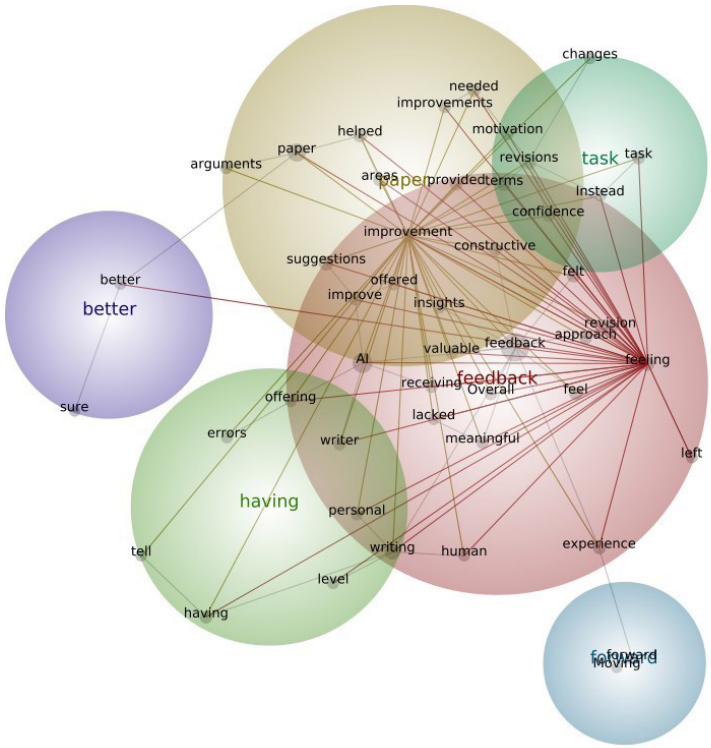


Figure 3 Pathways between ‘feeling’ and ‘improvement’

Participants expressed a range of feelings regarding their motivation after receiving feedback, reflecting a mix of reactions. Some participants described the motivational effect of the feedback as mild, suggesting that while it provided some encouragement, it was not particularly inspiring. Others were more enthusiastic, explaining that the feedback gave them a sense of purpose and transformed an otherwise monotonous task into something more meaningful. For instance, several participants noted that the feedback helped them view the revision process as a new challenge rather than a repetitive activity. The term ‘excitement’ was mentioned by some in relation to their motivation, hinting at a potential connection between positive emotional states and their willingness to engage with revisions.

The concept pathway map for ‘motivation’ (Figure 4) highlights its intersection with several related terms, including ‘provided’, ‘revision/s’, ‘needed’, ‘improvement/s’, ‘help/ed’, and ‘change/s’. These connections suggest that participants perceived motivation as being tied to the practical outcomes of feedback, such as identifying areas for improvement and providing clear next steps for revisions. Additionally, links to concepts like ‘[moving] forward/s’ and ‘level’ indicate that participants felt motivation contributed to raising the quality of their revisions, enabling them to produce work at a higher standard.

Interestingly, despite the similarly high levels of effect for motivation and engagement observed in the quantitative analysis, the interviews revealed limited overlap between these two concepts. Few participants explicitly linked motivation and engagement in their responses. Furthermore, engagement did not emerge as a distinct concept in Leximancer’s analysis. It is possible that

the algorithm conflated engagement with motivation, given their conceptual similarity and overlapping usage in the interviews. This suggests that while participants may have experienced both motivation and engagement, they tended to describe their experiences in terms of motivation more frequently, which may have influenced the thematic analysis.

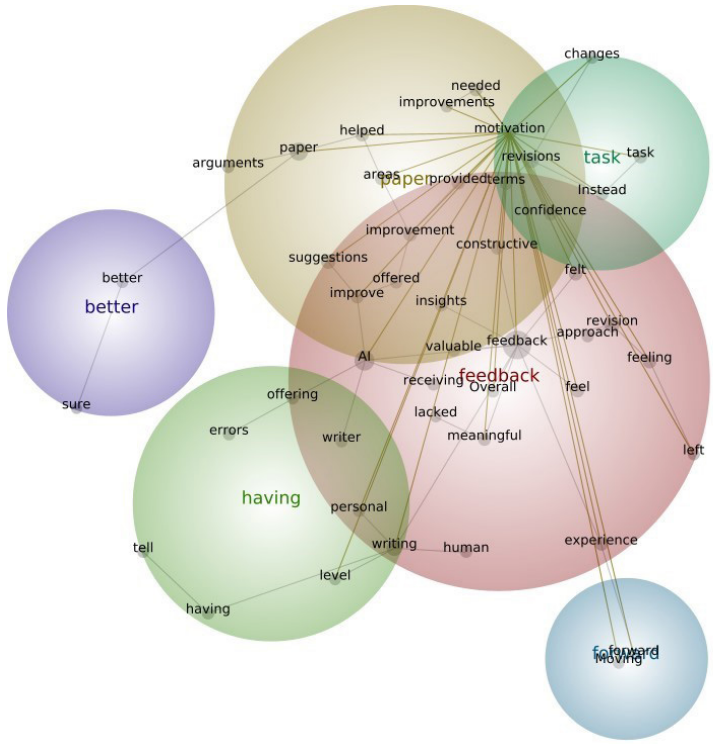


Figure 4 Pathways between 'motivation' and related concepts

Other themes identified in the analysis, such as 'having', appear to have been influenced by the phrasing of the interview prompts, as many participants began their responses with expressions like "it was like having". For example, participants described the experience of receiving AI feedback as being similar to "having a lecturer actually give feedback" or "having a second pair of eyes look at my paper".

The term 'paper' also emerged as a major concept in the analysis, frequently linked with 'motivation.' Participants often stated that receiving feedback from the AI increased their motivation, as it provided clear guidance on what they needed to address in order to improve their papers. The proximity of 'paper' and 'motivation' in the algorithmic analysis reflects this connection, suggesting that motivation was focused on completing revisions to their writing.

Several participants described the feedback as a kind of "roadmap" for improving their papers. They noted that the structured and detailed nature of the feedback helped them focus on specific goals during the revision process. For instance, some participants reported that knowing exactly what needed to be addressed allowed them to work efficiently and target specific

areas in their writing. The tailored nature of the feedback was also seen as a strength, with some expressing surprise at how well the AI identified issues in their papers. Emotional reactions to this specificity included a sense of being impressed or even shocked at the AI's ability to pinpoint relevant points for improvement.

The theme of 'task' was often associated with concepts such as 'insights,' 'improved,' and 'confidence,' indicating that participants found the feedback relevant to the task they were working on. Most participants agreed that the feedback was sufficiently task-specific, highlighting the utility of AI in responding to the content of a paper without requiring task-specific design, as is the case with Automated Writing Evaluations (AWEs). One participant even stated that they would use AI in the future to help revise essays prior to submission, citing its ability to improve final submissions beyond what could be achieved through proofreading alone. This view was shared across interviews, with no participants disputing the utility of the AI-generated feedback, although emotional and experiential responses to the feedback varied significantly.

When examining the theme of 'better,' both positive and negative associations were identified. Some participants acknowledged that the feedback helped them perform better on their assignments, although they still felt uneasy about receiving instruction from an AI. This discomfort was attributed to the impersonal nature of machine-generated feedback. Additionally, the analysis revealed that the term 'arguments' was associated with 'better' in 15% of cases, suggesting that participants felt the feedback helped them strengthen their arguments in the content of their papers. However, the absence of terms such as 'English' or 'language' in the analysis suggests that participants did not perceive the feedback as significantly improving the linguistic accuracy of their work, likely because the focus of the task was general rather than language-specific.

That said, some individual responses indicated that the AI-generated feedback did make corrections to their written English, with one participant noting that the feedback included specific suggestions for improving grammatical and linguistic accuracy. However, participants often distinguished between corrections to their English and actual improvements in their English skills. For example, one participant observed that while their essay's English was improved, they did not feel they had learned anything new about the English language because the feedback provided instructions rather than explanations.

Finally, while participants generally valued the feedback as a practical tool, many expressed a preference for feedback from teachers, lecturers, or other human markers. Participants described AI feedback as a useful alternative when human feedback was unavailable but emphasized that it could not replace the expertise and personal insights offered by teachers. The interviews suggest that this preference for human feedback was widely held, though the study did not directly compare AI-generated feedback with teacher feedback. As such, it remains unclear how accurate participants' assumptions about the superiority of human feedback are, though their responses reflect a general belief that human feedback is more effective and desirable.

DISCUSSION

The results above highlight several trends that merit further discussion. The quantitative analysis demonstrated that the feedback group achieved higher revised scores and reported greater levels of emotion, motivation, and engagement compared to the control group. T-tests showed that these differences were statistically significant, except for the differences in emotional responses to the feedback. This aligns with the findings from the point biserial correlation, which revealed a weaker relationship between feedback and positive emotions, as well as the Pearson's correlation coefficient, which indicated a relatively less pronounced connection between emotional responses and score improvements during revision.

The qualitative analysis further revealed mixed emotional responses to AI feedback. While some participants acknowledged positive aspects of the feedback, such as its utility and specificity, others expressed dissatisfaction, describing feelings of disengagement and lack of inspiration. This finding contrasts with previous studies, such as those by Li and Xing (2021), which demonstrated that LLM-based feedback can have a statistically significant and positive impact on students' affective experiences and emotional well-being. The discrepancy may be attributed to the inherently critical nature of feedback, which can elicit negative emotions regardless of the source. Furthermore, some participants perceived AI feedback as inherently inferior to human feedback, citing its lack of personalization and empathy as significant drawbacks.

It is also worth considering that the feedback process itself—whether AI-generated or human—can sometimes be emotionally taxing for students. Previous research by McGarrell and Verbeem (2007) noted that feedback cycles, particularly when iterative or critical, can lead to emotional exhaustion. This phenomenon may be even more pronounced in the case of LLM-based feedback, perhaps due to its perceived detachment or the specific design of the study, which could have amplified these effects.

Interestingly, this mixed emotional response to feedback was not reflected in participants' reports of motivation, which were generally positive. Many participants linked their motivation to the improvements they made during the revision process, aligning with the strong correlations identified in the Pearson's correlation coefficient between motivation and revision scores and engagement and revision scores. Previous studies have highlighted the influence of motivation on measured outcomes in writing tasks (Schrader & Kalyuga, 2020) and the role of instructor feedback in successfully enhancing student motivation (Lipnevich et al., 2021). The findings of this study suggest that LLM-based feedback not only significantly enhances motivation but may also account for a substantial proportion of the observed differences in test scores between the feedback and control groups.

In terms of the effect of AI-generated feedback on test scores, the analysis revealed a statistically significant correlation between receiving feedback and the difference in marks between the original and revised written products. A t-test comparing the control and feedback groups confirmed a statistically significant difference in score improvements, showing that participants who received feedback performed better on their revisions. Additionally, a point biserial correlation analysis indicated a weak positive effect of receiving feedback on outcomes.

These findings contribute to addressing the gap identified by Tao et al. (2024), who noted a lack of empirical evidence supporting the effect of LLM-based feedback on learning outcomes. While the effect size observed in this study was modest, the results provide evidence that AI-generated feedback can positively impact revision performance and learning outcomes, particularly by enhancing motivation.

INFERENCES, LIMITATIONS AND IMPLICATIONS

It is important to acknowledge that this study was conducted with first-year university students in Hong Kong, a specific educational and cultural context. As such, the findings may not fully generalize to students in different countries or educational systems, where factors such as language proficiency, learning styles, or attitudes toward AI may differ significantly. Future research could explore the applicability of AI-generated feedback in diverse contexts to better understand how local educational environments shape its effectiveness.

Based on the findings above, it appears that LLM-based feedback positively influences test scores by providing sufficiently targeted and actionable feedback. Even among students who reported negative impacts on their emotional state, the utility and focus of the feedback were not disputed, suggesting that its mechanism of effect operates more through motivation and engagement than through emotional shifts. A plausible explanation is that the key shift lies in participants' attitudinal disposition toward the task. Many interviewees noted that the feedback gave the task a sense of purpose or meaning, emphasizing the value of having specific, actionable guidance available during revision.

However, there is a possibility that these findings may not fully translate to real-world scenarios. The observed improvement in motivation may have been context-dependent, driven more by the feedback providing a sense of purpose for the immediate task than by the intrinsic utility of its guidance. Since the tasks in this study were not tied to course credits or qualifications, it is possible that revising the paper only felt worthwhile to participants because of the feedback and "instructions" provided by the AI.

A key limitation of this study is the lack of pre-feedback data. Without a pre-feedback questionnaire, it is not possible to compare participants' motivational or emotional states before and after receiving feedback. This limits the ability to measure how much the intervention influenced changes in these factors. Additionally, while the interview analysis provided valuable insights, the semantic focus of the algorithm used (Leximancer) did not fully capture the specific views, attitudes, and experiences of participants at an individual level. This restricts the depth of understanding regarding the relationships between their emotional or motivational states and their perceptions of the feedback.

Another limitation of this study concerns its applicability to domain-specific knowledge and subject-specific written tasks. The task assigned to the English-language students was relatively generic, lacking a clear connection to subject-specific skills or specialized content. As a result, participants in the interviews did not comment on whether the AI feedback improved their

writing skills. Instead, they primarily focused on how the feedback helped them generate new arguments or ideas. This observation aligns with concerns raised by Knoth et al. (2024), who highlight the limitations of AI in analytical tasks, suggesting it performs better in generating content than in providing deeper critical or analytical insights.

It is possible that the content of the feedback itself was not particularly helpful in improving students' writing and that the observed improvements in scores were primarily driven by the feedback's effect on motivation and engagement. The design of this study does not allow for an evaluation of how accurate, relevant, or helpful the AI's guidance was in terms of content or analysis. This remains an important area for further investigation.

Despite these limitations, the statistical analysis revealed that students who received AI-generated feedback experienced significant improvements in their scores compared to the control group. On average, the feedback group scored 3.113 marks higher on their revised papers, and statistical tests confirmed that these improvements were statistically significant. While the effect size of receiving AI feedback was weak ($r = 0.208$), the findings suggest a positive relationship between AI feedback and revision outcomes, even if the precise mechanism behind this relationship remains unclear. This highlights the need for future research that controls for additional variables to better understand the causal pathways involved.

The findings suggest that LLM-based generative AI has the potential to enhance the quality of students' written work. Students who received AI-generated feedback demonstrated measurable improvements in their revisions, as reflected in the quantitative analysis. Motivation and engagement were found to have strong positive correlations with score improvements, while the feedback group also showed a minor improvement in these measures relative to the control group. However, the relationship between emotional positivity and feedback was weaker, with only a moderate correlation between emotional positivity and revision scores, and no significant link between feedback and emotional positivity itself.

Qualitative insights from the interviews reflect a mixed emotional response to receiving AI feedback. While some participants appreciated the actionable and timely nature of the feedback, others found it impersonal and uninspiring. Statistical analysis further supported the lack of a strong connection between feedback and emotional positivity. However, relationships between feedback and motivation and engagement were statistically significant, albeit weak in strength. Motivation and engagement, in turn, were strongly correlated with overall revision performance across both groups. This suggests that AI feedback may have contributed to increased motivation and deeper engagement with the task by providing structured suggestions for improvement.

Nevertheless, it remains uncertain whether these effects would carry over to real-world scenarios outside of the study's controlled environment. In real-world contexts, factors such as time constraints, external pressures, or varying levels of support could influence the extent to which students benefit from AI feedback. Additionally, other latent variables may have played a role in the observed improvements, such as students' prior writing ability, their familiarity with receiving feedback, or their openness to using AI tools. Further research is

needed to explore these variables and to determine whether AI feedback can sustain its positive impact under less controlled conditions.

CONCLUSION

This study highlights the potential of AI-generated feedback to modestly improve the revision of written work, with slight increases in motivation and engagement appearing to play a role in this relationship. While the observed improvements were statistically significant, the effect size was relatively small, suggesting that AI feedback, while useful, may not yet match the depth and impact of human feedback in its current form. Nonetheless, these findings underscore the growing relevance of LLM-based AI tools in educational contexts and their potential to support students in revising and improving their work, particularly when human feedback is unavailable.

One of the key insights from this study is the variability in students' emotional responses to AI feedback. While some participants appreciated the actionable and targeted nature of the feedback, others reported feeling disengaged or uninspired. This variability raises important questions about the role of personalization in AI feedback and how it might influence emotional states, attitudinal shifts, and performance outcomes. Future research could delve deeper into these relationships, particularly by exploring how pre-existing attitudes towards AI—such as skepticism or openness—affect students' emotional and motivational responses to feedback. Understanding these dynamics could inform the development of more tailored and empathetic AI systems that better address individual needs and preferences.

A recurring theme in participants' responses was the perceived inferiority of AI feedback compared to human feedback, which was often attributed to the lack of personalization, empathy, and nuanced understanding that human instructors provide. This invites further research directly comparing the experience, accuracy, relevance, and outcomes of AI-generated feedback versus human feedback. Such studies could provide valuable insights into the contexts in which AI feedback is most effective and where it falls short, helping educators determine how best to integrate AI tools into existing feedback processes.

While the study demonstrates the potential of AI feedback in controlled experimental conditions, its applicability in real-world educational settings remains uncertain. Longitudinal studies that measure the sustained use of AI feedback in authentic classroom environments would be crucial in assessing its transferability to broader contexts. These studies could evaluate how AI feedback interacts with factors such as time constraints, external pressures, and individual differences in learning styles, as well as its effectiveness in fostering long-term skill development rather than short-term task completion.

Additionally, this study highlights the need to explore the effectiveness of AI feedback in domain-specific assessments. The task used in this study was relatively generic, and participants focused more on the AI's ability to generate new arguments or ideas rather than its impact on writing skills or subject-specific competencies. Future research could investigate how AI

feedback functions in more specialized contexts, such as scientific writing, creative writing, or technical domains, where its ability to provide accurate content-specific guidance may vary. Similarly, differentiating between the specific outcomes of AI feedback—such as improvements in writing mechanics or content development—could help educators understand the precise mechanisms through which AI impacts student performance.

Finally, this study suggests that AI feedback may hold particular promise for students who lack access to timely human feedback, such as those in large class settings or distance learning environments. By offering immediate and actionable suggestions, LLM-based feedback tools could help bridge gaps in access to feedback, particularly when human resources are limited. However, the findings also caution against over-reliance on AI tools, as they may not yet fully replicate the depth or emotional support provided by human instructors. While this study demonstrates the potential of AI-generated feedback to improve essay writing and foster motivation among first-year university students in Hong Kong, the findings should be interpreted within the specific cultural and educational context of the study. Further research is needed to investigate whether these results can be replicated in different countries, educational systems, or among students with varying levels of language proficiency and familiarity with AI tools.

In conclusion, while this study demonstrates the potential of AI-generated feedback to enhance the revision process, it also raises important questions about its limitations and the contexts in which it is most effective. To fully realize the potential of AI in education, future research should focus on addressing these limitations, refining the personalization and analytical capabilities of AI tools, and exploring their integration into diverse educational environments. While this study highlights the short-term benefits of AI-generated feedback for improving essay revisions, the long-term impact on students' overall writing skills remains unclear. Future research could explore whether these short-term improvements persist over time and contribute to sustained growth in academic writing proficiency. Longitudinal studies could examine how repeated exposure to AI feedback influences students' ability to self-regulate their writing and apply feedback independently in future tasks, providing deeper insights into the long-term educational value of such tools.

THE AUTHORS

Sumie Chan is currently teaching in the English Language Teaching Unit at The Chinese University of Hong Kong, China. She has 10 years of full-time teaching experience at various universities with a variety of English and literature courses, including English for academic purposes, English for science, English for arts, literary studies, visual culture, English for engineering, university English, Shakespeare and Hollywood cinema. Her research areas lie mainly in Shakespearean literature, performing arts and comparative literature.
sumiechan@cuhk.edu.hk

Noble Lo is currently pursuing his PhD in the Department of Educational Research at the Faculty of Arts and Social Sciences, Lancaster University, United Kingdom. He is also teaching at the College of Professional and Continuing Education (CPCE), Hong Kong Polytechnic University in Hong Kong. He has authored over 100 publications across diverse subject areas, including law, education, literature, and translation. His co-edited books include *Critical Reflections on ICT and Education* (Springer, 2023) and *Transcending Boundaries in the Digital Age* (Springer, 2025).
p.k.lo@lancaster.ac.uk

Alan Wong is a lecturer in the English Language Teaching Unit at The Chinese University of Hong Kong, China. He has presented at multiple international conferences and published papers on topics including the use of lesson recordings, student engagement, and shadow education. His research interests mainly include English language education, public speaking, and teacher language awareness.
awalanwong@cuhk.edu.hk

REFERENCES

Abendschein, B., Lin, X., Edwards, C., Edwards, A., & Rijhwani, V. (2024). Credibility and altered communication styles of AI graders in the classroom. *Journal of Computer Assisted Learning, 40*(4), 1766–1776. https://doi.org/10.1111/jcal.12979

Alvero, A. J., Arthurs, N., Antonio, A. L., Domingue, B. W., Gebre-Medhin, B., Giebel, S., & Stevens, M. L. (2020). AI and holistic review: Informing human reading in college admissions. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 20*, 200–206. https://doi.org/10.1145/3375627.3375871

Al Shloul, T., Mazhar, T., Abbas, Q., Iqbal, M., Ghadi, Y. Y., Shahzad, T., Mallek, F., & Hamam, H. (2024). Role of activity-based learning and ChatGPT on students' performance in education. *Computers and Education: Artificial Intelligence, 6*, Article 100219. https://doi.org/10.1016/j.caeai.2024.100219

Al-Khreseh, M. H. (2024). Bridging technology and pedagogy from a global lens: Teachers' perspectives on integrating ChatGPT in English language teaching. *Computers and Education: Artificial Intelligence, 6*, Article 100218. https://doi.org/10.1016/j.caeai.2024.100218

Aslan, S., Durham, L. M., Alyuz, N., Okur, E., Sharma, S., Savur, C., & Nachman, L. (2024). Immersive multi-modal pedagogical conversational artificial intelligence for early childhood education: An exploratory case study in the wild. *Computers and Education: Artificial Intelligence, 6*, Article 100220. https://doi.org/10.1016/j.caeai.2024.100220

Attride-Stirling, J. (2001). Thematic networks: An analytical tool for qualitative research. *Commission for Health Improvement, 1*(3), 385–405. https://doi.org/10.1177/146879410100100307

British Educational Research Association. (2018). *Ethical guidelines for educational research* (4th ed.).

Bowman, S. R. (2023, April 2). *Eight things to know about large language models*. arXiv. https://doi.org/10.48550/arXiv.2304.00612

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. http://dx.doi.org/10.1191/1478088706qp063oa

Bressane, A., Zwirn, D., Essiptchouk, A., Saraiva, A. C. V., de Campos Carvalho, F. L., Formiga, J. K. S., de Castro Medeiros, L. C., & Negri, R. G. (2024). Understanding the role of study strategies and learning disabilities on student academic performance to enhance educational approaches: A proposal using artificial intelligence. *Computers and Education: Artificial Intelligence, 6*, Article 100196. https://doi.org/10.1016/j.caeai.2023.100196

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chen, S. M., & Bai, S. M. (2010). Using data mining techniques to automatically construct concept maps for adaptive learning systems. *Expert Systems with Applications, 37*(6), 4496–4503. https://doi.org/10.1016/j.eswa.2009.12.060

Cheng, C. (2024). Using AI-generative tools in tertiary education: Reflections on their effectiveness in improving tertiary students' English writing abilities. *Online Learning, 28*(3), 33–54. https://doi.org/10.24059/olj.v28i3.4632.

Chia, Y. K., Hong, P., Bing, L., & Poria, S. (2023, June 15). *INSTRUCTEVAL: Towards holistic evaluation of instruction-tuned large langauge models*. arXiv. https://doi.org/10.48550/arXiv.2306.04757

Crossley, S. A., Baffour, P., Tian, Y., Picou, A., Banner, M., & Boser, U. (2022). The persuasive essays for rating, selecting, and understanding argumentative and discourse element (PERSUADE) corpus 1.0. *Assessing Writing, 54*, Article 100667. https://doi.org/10.1016/j.asw.2022.100667

Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y. S., Gasevic, D., & Chen, G. (2023). Can large language models provide feedback to students? A case study on ChatGPT. *IEEE International Conference on Advanced Learning Technologies (ICALT), 2023*, 323–325. https://doi.org/10.1109/ICALT58122.2023.00100

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology, 61*, Article 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Essel, H. B., Vlachopoulos, D., Essuman, A. B., & Amankwa, J. O. (2024). ChatGPT effects on cognitive skills of undergraduate students: Receiving instant responses from AI-based conversational large language models (LLMs). *Computers and Education: Artificial Intelligence, 6*, Article 100198. https://doi.org/10.1016/j.caeai.2023.100198

Feng, S., & Law, N. (2021). Mapping artificial intelligence in education research: A network-based keyword analysis. *International Journal of Artificial Intelligence in Education, 31*, 277–303. https://doi.org/10.1007/s40593-021-00244-4

Fleckenstein, J., Liebenow, L. W., & Meyer, J. (2023). Automated feedback and writing: A multi-level meta-analysis of effects on students' performance. *Frontiers in Artificial Intelligence, 6*, Article 1162454. https://doi.org/10.3389/frai.2023.1162454

Gao, R., Merzdorf, H. E., Anwar, S., Hipwell, M. C., & Srinivasa, A. (2024). Automatic assessment of text-based responses in post-secondary education. *Computers and Education: Artificial Intelligence, 6*, Article 100206. https://doi.org/10.1016/j.caeai.2024.100206

Gnepp, J., Klayman, J., Williamson, I. O., & Barlas, S. (2020). The future of feedback: Motivating performance improvement through future-focused feedback. *PLoS One, 15*(6), Article e0234444. https://doi.org/10.1371%2Fjournal.pone.0234444

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing. *The Elementary School Journal, 115*(4), 523–547. https://doi.org/10.1086/681947

Hahn, M. G., Navarro, S. M. B., Valentin, L. D. L. F., & Burgos, D. (2021). A systematic review of the effects of automatic scoring and automatic feedback in educational settings. *IEEE Access, 9*, 108190–108198. https://doi.org/10.1109/ACCESS.2021.3100890

Holmes, A. G. D. (2020). Researcher positionality – A consideration of its influence and place in qualitative research – A new researcher guide. *Shanlax International Journal of Education, 8*(4), 1–10. http://dx.doi.org/10.34293/education.v8i4.3232

Huang, A. Y., Lu, O. H., & Yang, S. J. (2023). Effects of artificial intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers & Education, 194*, Article 104684. https://doi.org/10.1016/j.compedu.2022.104684

Hwang, G. J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 1*, Article 100001. https://doi.org/10.1016/j.caeai.2020.100001

Jacobsen, L. J., & Weber, K. E. (2023). *The promises and pitfalls of ChatGPT as a feedback provider in higher education: An exploratory study of prompt engineering and the quality of AI-driven feedback*. OSF Preprints. http://dx.doi.org/10.31219/osf.io/cr257

Kabudi, T., Pappas, I., & Olsen, D. H. (2021). AI-enabled adaptive learning systems: A systematic mapping of the literature. *Computers & Education: Artificial Intelligence, 2*, Article 100017. https://doi.org/10.1016/j.caeai.2021.100017

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T. … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 6*, Article 100225. https://doi.org/10.1016/j.caeai.2024.100225

Langley, P. (2019). An integrative framework for artificial intelligence. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(1), 9670–9677. https://doi.org/10.1609/aaai.v33i01.33019670

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strwlan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024a). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence, 6*, Article 100221. https://doi.org/10.1016/j.caeai.2024.100221

Lee, G. G., Latif, E., Wu, X., Liu, N., & Zhai, X. (2024b). Applying large language models and chain-of-thought for automatic scoring. *Computers and Education: Artificial Intelligence, 6*, Article 100213. https://doi.org/10.1016/j.caeai.2024.100213

Lee, S., & Moore, R. (2024). Harnessing generative AI (GenAI) for automated feedback in higher education: A systematic review. *Online Learning, 28*(3), 82–106. https://doi.org/10.24059/olj.v28i3.4593

Li, C., & Xing, W. (2021). Natural language generation using deep learning to support MOOC learners. *International Journal of Artificial Intelligence in Education, 31*, 186–214. https://doi.org/10.1007/s40593-020-00235-x

Lipnevich, A. A., Murano, D., Krannich, M., & Goetz, T. (2021). Should I grade or should I comment: Links among feedback, emotions, and performance. *Learning and Individual Differences, 89*, Article 102020. https://doi.org/10.1016/j.lindif.2021.102020

Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour, 1*, Article 0028. https://doi.org/10.1038/s41562-016-0028

Madigan, D., & Kim, L. (2021). Does teacher burnout affect students? A systematic review of its association with academic achievement and student-reported outcomes. *International Journal of Educational Research, 105*, Article 101714. https://doi.org/10.1016/j.ijer.2020.101714

Magaldi, D., & Berler, M. (2020). Semi-structured interviews. In V. Zeigler-Hill & T. Shackelford (Eds.), *Encyclopedia of personality and individual differences* (pp. 4825–4830). Springer.

McCormick, K., & Salcedo, J. (2015). *SPSS statistics for dummies*. John Wiley.

McGarrell, H., & Verbeem, J. (2007). Motivating revision of drafts through formative feedback. *ELT Journal, 61*(3), 228–236. https://doi.org/10.1093/elt/ccm030

Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology, 114*(8), 1743–1772. http://dx.doi.org/10.1037/edu0000764

Meyer, J., Jansen, T., Schiller, R., Liebenow, L. W., Steinbach, M., Horbach, A., & Fleckenstein, J. (2024). Using LLMs to bring evidence-based feedback into the classroom: AI-generated feedback increases secondary students' text revision, motivation, and positive emotions. *Computers and Education: Artificial Intelligence, 6*, Article 100199. https://doi.org/10.1016/j.caeai.2023.100199

Misiejuk, K., Kalissa, R., & Scianna, J. (2024). Augmenting assessment with AI coding of online student discourse. *Computers and Education: Artificial Intelligence, 6*, Article 100216. https://doi.org/10.1016/j.caeai.2024.100216

Pandero, E., & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, Article 100416. https://doi.org/10.1016/j.edurev.2021.100416

Peters, K., & Halcomb, E. (2015). *Interviews in qualitative research. Nurse Researcher, 22*(4), 6–7. https://doi.org/10.7748/nr.22.4.6.s2

Ramesh, D., & Sanampudi, S. K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review, 55*, 2495–2527. https://doi.org/10.1007/s10462-021-10068-2

Salcedo, J., & McCormick, K. (2020). *SPSS statistics for dummies* (4th ed.). John Wiley.

Saúde, S., Barros, J. P., & Almeida, I. (2024). Impacts of generative artificial intelligence in higher education: Research trends and students' perceptions. *Social Sciences, 13*(8), Article 410. https://doi.org/10.3390/socsci13080410

Schrader, C., & Kalyuga, S. (2020). Linking students' emotions to engagement and writing performance when learning Japanese letters with a pen-based tablet: An investigation based on individual pen pressure parameters. *International Journal of Human-Computer Studies, 135*, Article 102374. https://doi.org/10.1016/j.ijhcs.2019.102374

Smith, A. E., & Humphreys, M. S. (2006). Evaluation of unsupervised semantic mapping of natural. *Behaviour Research Methods, 38*(2), 262–279. https://doi.org/10.3758/BF03192778

Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., Moon, Y., Tseng, W., Waschauer, M., & Olsen, C. B. (2024). Comparing the quality of human and ChatGPT feedback on students' writing. *Learning and Instruction, 91*, Article 101894. https://doi.org/10.1016/j.learninstruc.2024.101894

Tao, Z., Lin, T., Chen, X., Li, H., Wu, Y., Li, Y., Jin, Z., Huang, F., Tao, D., & Zhou, J. (2024). *A survey on self-evolution of large language models*. ArXiv. https://doi.org/10.48550/arXiv.2404.14387

Wang, D. (2024). Teacher- versus AI-generated (Poe Application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *The International Review of Research in Open and Distributed Learning, 25*(3), 37–56. https://doi.org/10.19173/irrodl.v25i3.7646

Wardat, Y., Tashtoush, M. A., AlAli, R., & Jarrah, A. M. (2023). ChatGPT: A revolutionary tool for teaching and learning mathematics. *EURASIA Journal of Mathematics, Science and Technology Education, 19*(7), Article em2286. https://doi.org/10.29333/ejmste/13272

Weber, F., Wambsganss, T., & Söllner, M. (2024). Enhancing legal writing skills: The impact of formative feedback in a hybrid intelligence learning environment. *British Journal of Educational Technology*. Advance online publication. https://doi.org/10.1111/bjet.13529

Yang, S., Nachum, O., Du, Y., Wei, J., Abbeel, P., & Schuurmans, D. (2023). *Foundation models for decision making: Problems, methods, and opportunities*. arXiv. https://doi.org/10.48550/arXiv.2303.04129

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0

Zheng, Y. D., & Stewart, N. (2024). Improving EFL students' cultural awareness: Reframing moral dilemmatic stories with ChatGPT. *Computers and Education: Artificial Intelligence, 6*, Article 100223. https://doi.org/10.1016/j.caeai.2024.100223

Appendix 1

Example of student essay, AI-generated feedback, and revise essay

Original Essay

Children in the 21st Century are using tablet computers and smart phones more than ever. A survey of 773 parents in Hong Kong now showed that children are 4 to 6 hours every day on electronic devices. Some suggest that this bad for children for reasons such as damage to eyesight, impaired gross motor function, emotional and well-being, and problems with social development. There are now some that say that children should not be allowed to use devices at all under a certain age because of the risks of damage associated with using electronic devices at a young age.

One of the main reasons that this might be a good idea is because children's brains are not formed by the age of 5. Young children's brains are still growing and have not become mature enough to cope with long use of electronic devices. It is argued that children require social action to be develop fully and that using devices all day can cause less time for social action. What's more, smartphones and computers can encourage the wrong type of social action, such as contacting adults on line or seeing adult content. How device use impacts children's health and brains is not well understood scientifically and therefore it is risk to allow children to use so much.

However, there are other arguments to say that using electronic devices is good or okay. For instance, there is more and more work at school from the primary level. Students in the future will probably have to complete work on computers or tablets more than ever, meaning students that don't use these will be left behind in adult world or work. More than ever, student social action takes place on line, meaning that students who don't have phones might be left out socially, hurting their developing. So there are a number of reasons why young children might benefit from using electronic devices.

My opinion is that it is up to parents if young children use devices. Government should guide parents but not tell them how to parent in law. This is part of ensuring freedom in Hong Kong and the world over. Parents should be told about the risks and benefits to young children using phones and tablets through public awareness but should not be made illegal for them to allow their children to use it. This is a fair way of doing things as there is no clear evidence either way as to what young children using electronic devices does.

AI Feedback (ChatGPT)

This essay addresses an important contemporary issue and presents arguments both for and against the prohibition of young children using electronic devices. While there are some strong points made, improvements are needed in critical engagement, clarity of argument, academic style, and grammatical accuracy in order to raise the overall quality of the work.

1. Engage Critically with the Question

The essay touches on both sides of the debate but lacks depth in critical engagement. You briefly mention potential risks (damage to eyesight, social development) and benefits (preparing children for future work, social inclusion), but these points are underdeveloped.

Suggestions for Improvement:

- Expand on the evidence supporting your claims. For instance, you mention risks such as impaired motor function or social development but do not provide any specific studies or research to support these ideas. Look for scientific studies or expert opinions that discuss these risks in detail.

- Critically evaluate counterarguments rather than just stating them. For example, you could discuss whether preparing children for a digital future outweighs the health risks or explore if limited, supervised use might be a compromise.

- Provide more specific examples, especially for the opposing viewpoint. What kinds of positive educational outcomes might children experience by using tablets or smartphones?

2. Argue Convincingly Towards a Clear Thesis

Your thesis—expressing that parents should decide on the use of devices rather than the government—is introduced at the end but lacks a clear presence throughout the essay. Additionally, the essay feels somewhat unbalanced, with the pros and cons not clearly tied to your overall argument.

Suggestions for Improvement:

- Clarify your thesis in the introduction so that readers know from the outset what your position is.

- Strengthen the coherence of your argument by making sure each paragraph relates to your thesis. For example, the second paragraph could more clearly connect the risks of device use to why you believe the decision should be left to parents.

- Transition between arguments and counterarguments more fluidly. Currently, the sections feel somewhat disconnected, and the essay would benefit from smoother progression.

3. Employ Academic Vocabulary and Formal English

The essay employs some basic academic vocabulary but could benefit from a more formal tone and a wider range of advanced language. Informal phrases like "what's more" and "okay" are more suited to conversational English than academic writing.

Suggestions for Improvement:

- Replace informal phrases with more formal equivalents. For example, instead of "what's more", use "additionally" or "furthermore".

- Use more sophisticated academic language to express complex ideas. For instance, "it is argued" is a good start, but phrases like "it has been suggested by research" or "empirical studies indicate" would strengthen your academic tone.

- Avoid using contractions like "don't" and "it's" in formal writing. Replace them with "do not" and "it is".

4. Ensure Accurate Spelling and Grammatical Coherence

There are several grammatical issues in the essay, particularly with sentence structure and word choice. Errors such as "are 4 to 6 hours every day on electronic devices" and "social action to be develop fully" detract from the clarity and flow of your ideas.

Suggestions for Improvement:

- Pay attention to subject-verb agreement and verb forms. For instance, "are 4 to 6 hours" should be "spend 4 to 6 hours".

- Review your use of articles (e.g., "a risk to allow children to use so much" could be "it is risky to allow children to use them so often").

- Proofread for spelling errors and awkward phrasing. For example, "emotion and well-being" should be "emotional well-being", and "students that don't use these will be left behind in adult world or work" could be rephrased as "students who do not use these devices may struggle in the adult world or workforce".

General Guidance

- Research and Evidence: Incorporate more references to existing studies or expert opinions. This will support your claims and improve the persuasiveness of your argument.

- Structure: Organize your essay more clearly by using topic sentences at the start of each paragraph to signal the argument. Make sure every paragraph ties back to your main thesis.

- Language and Style: Aim to elevate your use of academic vocabulary and avoid informal expressions to maintain a professional tone throughout.

Conclusion

By expanding your arguments, improving grammatical accuracy, and adopting a more academic tone, you can significantly improve the quality of your essay. Clear evidence, stronger engagement with the question, and consistent focus on your thesis will help to create a more convincing and coherent response.

Revised Essay

Children in the 21st Century are using tablet computers and smart phones more than ever. A survey of 773 parents in Hong Kong now showed that children spend 4 to 6 hours every day on electronic devices (Wu, 2017). It has been suggested by research that this bad for children for reasons such as damage to eyesight, gross motor function, emotional well-being, and social development (The Economist, 2024). There are now some experts that say that children should not be allowed to use devices at all under a certain age because of the risks of damage associated with using electronic devices at a young age (Addley, 2023). This essay argues that because of the conflicting evidence on the effects of device use in young children, government should guide parents but not ban use of electronic devices by children under 5.

One of the main reasons that banning devices might be a good idea is because empirical studies indicate children's brains are not formed by the age of 5 (Mojica, 2023). Young children's brains are still growing and have not become mature enough to cope with long use of electronic devices. Furthermore, it is argued that children require social interaction to develop fully and that using devices all day can cause less time for social interaction. Studies show that children rely on play to develop emotionally and that it is good for their well-being also (Ng, 2023). Additionally, smartphones and computers can encourage the wrong type of social interaction, such as contacting adults online or seeing adult content. There may be physical problems as well, such as affecting eyesight and gross motor function. For example, studies show that during Covid eyesight among young children worsened in Hong Kong (Ho, 2023). However, there are experts that disagree with these claims and argue that the evidence for long term effects are not established (NCT).

There are other arguments to say that using electronic devices is acceptable or even good for children. For instance, there is more and more work at school from the primary level that is on devices. Students in the future will probably have to complete work on computers or tablets more than ever, meaning students that do not use these devices may struggle in the adult world or workforce. More than ever, student social interaction takes place on line, meaning that students who do not have phones might be left out socially, hurting their development (Bozzola, 2022). So there are a number of reasons why young children might benefit from using electronic devices. These reasons need to be critically evaluated so better arrive at a compromise for device use.

Children must be prepared for a digital future as that becomes reality. However, there are also reasons why this is not the most important concern. If children's emotional well-being, social development and physical health is risked, then this digital education can wait until they are

more developed. However, the disagreed evidence above indicates that there is not a clear effect on children by early device use. Making this illegal could make parents criminals for letting their children use their phone or watch television on a device. This damages freedom for all and means that government might go too far beyond what evidence suggests is necessary.

In conclusion, government should guide parents but not tell them how to parent in law. Parents should be told about the risks and benefits to young children using phones and tablets through public awareness but it should not be made illegal for them to allow their children to use it. This is a fair way of doing things as there is no clear evidence either way as to what young children using electronic devices will do to them.

Appendix 2

Post-experiment questionnaire delivered to participants

Post-Experimental Questionnaire

Thank you for participating in this research. Please take the time to answer a few questions about your experience. Your responses will help the researchers better understand how students engage with revising their work.

Below is a series of statements about your experience and attitudes towards the revision undertaken and we ask that you indicate the extent to which you agree or disagree with each statement. Please rate each statement on a scale from 1 to 10 (where 1 is strongly disagree and 10 is strongly agree) by circling the appropriate number corresponding with your outlook.

Your responses will be anonymised prior to analysis and will be used solely for the purposes of statistical research, so please answer as honestly as possible.

1. I felt positive about the process of making revisions to my work.

1 2 3 4 5 6 7 8 9 10

2. I was motivated to complete revisions on my work.

1 2 3 4 5 6 7 8 9 10

3. I found the process of completing revisions to be satisfying.

1 2 3 4 5 6 7 8 9 10

4. The process of revising my work was engaging.

1 2 3 4 5 6 7 8 9 10

5. I felt confident in my ability to make revisions that would improve my work.

1 2 3 4 5 6 7 8 9 10

6. I found it easy to concentrate when revising my work.

1 2 3 4 5 6 7 8 9 10

7. I knew which areas of my work would benefit from revision.

1 2 3 4 5 6 7 8 9 10

8. I felt that the time put into revising my work was worthwhile.

1 2 3 4 5 6 7 8 9 10

9. The feedback I received helped me feel more motivated to revise my work.

1 2 3 4 5 6 7 8 9 10

10. I enjoyed the process of revising my work.

1 2 3 4 5 6 7 8 9 10