From Uniform Feedback to Adaptive Support: AI-Driven Personalized Writing Feedback for Enhancing Self-Assessment Accuracy in Higher Education

我這邊就同時把reviewer意見，要修改的內容，以及回復review的話都先寫出來了。那先給你過目，OK的話，我再用原本的點，去多複製出其他reviewer的意見。

**Reviewer 1:**

* Abstrac第一段 “With the rapid advancement of generative artificial intelligence, large language models (LLMs) have become increasingly integrated into education, particularly for automated formative feedback and writing assessment.” 該段內容較多著重背景敘述，未能明確呈現本研究的核心貢獻與角色。

感謝委員指出 Abstract 第一段過於著重背景敘述，未能充分凸顯本研究之目的與貢獻。我們依據建議，已修訂該段內容，刪減原本偏向背景鋪陳的句子，並重新聚焦於研究動機與學術貢獻。修訂後的內容如下：

“With the growing integration of large language models (LLMs) into education, their use in formative feedback and writing assessment has attracted increasing attention. This study explores how such technologies can enhance students’ self-assessment accuracy in higher education writing.”

此修改簡化了句構、突顯研究切入點，並清楚說明本研究的目的與角色。我們再次感謝委員的寶貴建議。

* Abstract第二段 “This study introduces and evaluates an AI-driven intelligent feedback system designed to enhance sustainable and inclusive higher education, leveraging transformer-based models (BERT and RoBERTa) to provide scalable, adaptive, and personalized writing support. The system aims to improve students’ self-assessment accuracy (SAA), a critical factor for self-regulated learning, while addressing the challenge of delivering high-quality feedback efficiently in under-resourced contexts.” 該段同時呈現系統目的、技術細節、教育價值與應用情境，導致資訊密度較高，影響可讀性。建議將其拆分為兩句：第一句聚焦於系統目的與教育背景，第二句介紹技術與特性。

感謝委員指出論述過於負擔而導致承重的問題，這邊依據委員給予的建議，將原文中系統目的與技術細節分開敘述，以提升可讀性與結構清晰度。修改過後論述如下：

“This study introduces and evaluates an AI-driven feedback system aimed at improving self-assessment accuracy (SAA) in higher education writing, a key factor for self-regulated learning. The system leverages transformer-based models (BERT and RoBERTa) to deliver scalable, adaptive, and personalized writing support, addressing the challenge of providing high-quality feedback in under-resourced contexts.”

改後的版本將教育背景與系統目標獨立成首句，使讀者先理解研究目的與核心價值，再於第二句介紹系統技術與應用特性。這樣的安排可以降低資訊密度、提升閱讀流暢度，並凸顯系統的技術優勢與教育應用價值。

* Introduction第四段 ” Nonetheless, evidence regarding the effectiveness of LLM-generated feedback in supporting SAA remains inconclusive…” 這一段屬於研究缺口，但在論述上 “concerns remain” 、 “may fall short”用詞不夠具體有力。建議強化該段的論述脈絡，將「低表現學生需要明確標準」與「AI 回饋的潛力及限制」進行清晰串聯，以凸顯研究缺口的邏輯連結。同時，將 “concerns remain” 與 “may fall short” 改為更具體且有說服力的表述。

感謝委員指出該段落論述力度不足。我們已依據建議，將原本相對籠統的表述改為更具體且具說服力的詞彙，並加強邏輯串聯。修訂後版本如下：

“However, the effectiveness of LLM-generated feedback for improving SAA remains uncertain. While LLMs can provide rapid and consistent responses (Meyer et al., 2024), they may lack contextual sensitivity and fail to offer the cognitive scaffolding needed to guide strategic thinking or address nuanced errors. Such limitations could reduce their impact on students’ ability to recalibrate self-assessments. Prior research highlights that low-performing students particularly benefit from explicit performance standards and structured feedback (Lew et al., 2010), underscoring the need to ex-amine whether LLMs can meet these requirements.”

此版本明確將「低表現學生需要明確標準」與「AI 回饋的潛力與限制」串聯，使研究缺口的論述更為清晰且具說服力。

* Related Works 2.1開頭 “Empirical Foundations of Self-Assessment Accuracy and Feedback in Scalable Learning Contexts” 連續引用三個同義詞（calibration accuracy, metacognitive monitoring accuracy），雖然完整，但對首次閱讀的讀者來說資訊密度過高，應該先給最常用的定義，再提其他術語作補充。

感謝委員的提醒，Related Works 2.1段落開頭連續引用三個同義詞（calibration accuracy、metacognitive monitoring accuracy）可能導致資訊密度過高，對首次閱讀的讀者不夠友善。我們依據建議調整了敘述順序，先提供最常用的定義與名稱，再補充其他學術用語作為說明，以提升可讀性並促進循序理解。修訂後版本如下：

” Self-assessment accuracy (SAA) refers to the degree of alignment between learners’ self-evaluations and their actual performance (Hacker & Bol, 2019). Also known as calibration accuracy or metacognitive monitoring accuracy (de Bruin and van Merriënboer, 2017).”

* Related Works 2.2第一段對Feedback literacy 的功能性描述重複出現(e.g. engage in meaningful self-assessment and benefit from it*、*complements self-assessment)，導致論述顯得冗贅，建議刪減重複的表述，直接切入定義主題，讓讀者更快理解核心概念。此外，後續論述“Subsequent research further identifies three core dimensions…” 、 ” Nicol (2021) introduces the concept of internal feedback…”兩者之間關係不夠明顯。建議強化兩者之間的過渡。

感謝委員指出 Related Works 2.2 開頭有功能性描述重複及核心構面與內在回饋之間過渡不足的情況。我們依據建議進行了修訂：

“Feedback literacy refers to the understandings, skills, and dispositions that enable learners to interpret, evaluate, and use feedback to improve performance (Carless & Boud, 2018; Molloy et al., 2020). Its core dimensions include proactively seeking relevant information, processing feedback effectively, and taking informed action (Malecka et al., 2020). Nicol (2021) extends this concept by introducing the notion of internal feedback, which involves self-generated cognitive evaluations that complement external input and contribute to deepening comprehension and application. In the context of AI-supported learning, these dimensions form the basis for learners’ capacity to integrate human and machine feedback into their self-regulation processes.”。

此修訂刪減了開頭冗餘的功能性表述，直接引入 Feedback literacy 的定義，使讀者能更快掌握核心概念，並加強了段落間的銜接性。

* 段落3.1 “Dataset and Preprocessing” 雖然數據來源交代完整，但在 “These texts include argumentative and narrative essays…” 的描述過於概略。建議明確說明 argumentative essays 與 narrative essays 的數量。此外，後續論述 “authentic educational context” 缺乏實證支撐，建議刪減或補充說明其可被視為 “authentic” 的依據。

感謝委員的寶貴建議，已依意見進行以下修正：

1. 在描述資料集時，補充了 argumentative essays 與 narrative essays 的具體數量，避免過於概括的表述。
2. 將原本「authentic educational context」的敘述改為具體說明資料收集於課堂評量中且無外部協助，藉此提供可視為「authentic」的實證依據。修訂後版本如下：

“The dataset consists of 7,158 essays written by undergraduate students enrolled in Mandarin courses at the university’s General Education Center between January 2024 and July 2025. The corpus includes 3,876 argumentative essays and 3,282 narrative essays, collected as part of regular course assessments without external assistance, thereby reflecting students’ natural Chinese writing performance in a real classroom context.”

* 段落4.1 “Study Sample” 雖然已提供參與者的人數、性別比例、年齡與分組方式，但未交代其 inclusion criteria、exclusion criteria，以及選擇工程學院學生作為樣本的理由。建議補充相關資訊，以提升研究方法的透明度與說服力。

感謝委員的建議。我們已依據建議補充了樣本的 inclusion criteria、exclusion criteria，以及選擇工程學院學生作為樣本的理由，以提升研究方法的透明度與完整性。修訂後版本如下：

“The participants in this study consisted of 64 undergraduate students (N = 64) enrolled in the College of Engineering at a university in southern Taiwan. Data collection was conducted between April and July 2024. The experimental course was part of the university’s general education program, focusing on the development of humanities literacy. Inclusion criteria required students to be officially enrolled in the targeted general education course during the study period, aged between 18 and 20 years, and with Mandarin as their primary language of instruction. Exclusion criteria included prior participation in similar training programs, absence from more than 20% of class sessions, or failure to complete major course assignments. The College of Engineering was selected as the sample source because all students in this college were required to take the course during the specified semester, ensuring consistency in baseline academic background and facilitating controlled comparisons. To ensure baseline equivalence in academic background and foundational abilities, students were assigned to either the experimental group (EG, N = 32) or the control group (CG, N = 32) using a stratified random assignment procedure. Among the participants, 56.75% were male with a mean age of 18.2 years, while 44.25% were female with a mean age of 18.7 years.”

* 段落5.1 “Descriptive Statistics”中論述 ” may provide meaningful support” 這句措辭偏口語化，建議在 Results 部分保持客觀描述，以維持結果呈現的中立性並避免過度推論。

感謝委員的提醒。我們已將原句 "may provide meaningful support" 修改為客觀的統計描述，以避免在 Results 中出現推論性語句，並回應委員對保持結果中立性的建議。修訂後版本如下：

“LLMF exhibited moderate positive partial correlations with both SAA (r = 0.44, p < .01\*\*) and LP (r = 0.36, p < .01\*\*). The highest partial correlation was observed between SAA and LP (r = 0.51, p < .01\*\*), revealing a stable and positive relationship between accurate self-assessment and enhanced learning performance.”

* 段落Discussion最後一段尾句 “should avoid a one-size-fits-all approach”， 雖然該表述有效傳達了不宜採用單一回饋模式的觀點，但內容偏概略，缺乏具體的系統設計建議，可能導致讀者難以掌握其在實務上的應用方式。建議補充更明確且可操作的改進策略，以增強建議的實用性與說服力。

感謝委員的寶貴建議，我們已經將原先較籠統的論述，修改為更具體的系統設計策略說明。修訂後版本如下：

“Therefore, implementing LLM feedback mechanisms should move beyond a one-size-fits-all approach by incorporating adaptive algorithms that adjust feedback depth, tone, and scaffolding level based on learners’ initial self-assessment accuracy and feedback literacy. Such differentiated and adaptive feedback strategies can maximize the educational potential of LLMs.”

在修訂版本中，我們補充了引入自適應演算法的設計構想，並說明可依據學習者的initial self-assessment accuracy 與 feedback literacy，動態調整回饋的深度、語氣與鷹架程度，以提高系統的針對性與成效，並回應委員對具體化改進策略的建議。