

通過人工智慧輔助同伴評價提高中學生學生寫作水準：與傳統同伴評價方法的比較研究

Enhancing Secondary School Students' Writing through AI-assisted Peer Evaluation: A Comparative Study with Traditional Peer Review Methods

謝施藝¹, 黃碧雲^{1*}, 鄭琨涵^{1,2}

¹School of Education, City University of Macau

²Guangzhou College of Applied Science and Technology

* byhuang@cityu.edu.mo

【摘要】本研究提出了一種人工智慧輔助同伴評價的方法，旨在提高初中生的英語寫作能力。本研究共招募 27 名初中學生作為研究對象，通過對比實驗組與對照組，發現人工智慧輔助同伴評價顯著提高了學生的英語寫作水準。實驗組學生接受教師、傳統同伴和人工智慧輔助同伴評價，而對照組學生只接受教師和同伴評價。研究結果表明，實驗組在寫作水準和自我效能感方面均優於接受傳統同伴評價的對照組，且寫作水準明顯提高，其自我效能感亦呈現積極變化。

【关键词】人工智慧輔助同伴評價；傳統互評；K-12；寫作成績

Abstract: In this study, an AI-assisted peer evaluation method is proposed to enhance junior high school students' English writing skills. A total of 27 junior high students were recruited to an experimental group and a control group. By comparing the two groups, it was found that AI-assisted peer evaluation significantly improved students' English writing ability. The experimental group received teacher evaluation, traditional peer evaluation, and AI peer evaluation, while the control group were only evaluated by the teacher and peers. The findings showed that the experimental group outperformed the control group in both writing proficiency and self-efficacy, showing significant improvement in writing and positive changes in self-efficacy.

Keywords: AI peer evaluation, traditional peer evaluation, K-12, writing performance

1. 前言

1.1 研究背景

人工智慧（AI）技術在教育領域的應用日益廣泛，尤其是在寫作評價和反饋方面(Gao & Merzdorf, 2024)。目前，初中生的寫作教學面臨著傳統教學方法難以充分滿足學生需求的挑戰。傳統教師評價常因反饋不及時、主觀性強等問題而受限。儘管同伴評價已被廣泛證明能夠有效提升學生的寫作能力（Hsia, L.-H., et al., 2015）。但其效果亦受限於學生的主觀性、知識水準和寫作能力以及評價經驗的不足（Guo et al., 2024）。

因此，引入人工智慧輔助同伴評價，透過智慧演算法提供一致性更高的評價標準，減少人為因素的干擾，提高評價的準確性與效率，為學生的寫作學習提供有力支援。

1.2 研究目的與問題

本研究聚焦於探討人工智慧輔助同伴評價相對於傳統同伴評價，在提高初中生寫作能力、影響學生自我效能感的效果。因此，本研究的研究問題如下：

- (1) 人工智慧輔助同伴評價與傳統互評哪個對促進學生英語寫作更有效？
- (2) 人工智慧輔助同伴評價和傳統互評對學生的自我效能感有何影響？

2. 文獻綜述

2.1 同伴評價在寫作中的效果與影響

同伴評價作為一種有效的學習策略，在寫作教學中尤其重要，學生們通過結對或小組合作，閱讀彼此的作文並提供建議，同時增強批判性思維及自我反思能力（Chang, 2016）。研究表明，同伴評價有助於改善文章的內容、結構和語言表達（昂娟&陳媛, 2024），減輕學生的焦慮，增強其完成任務的信心（Tseng, 2010）。此外，Zhao 等（2024）指出，書面對話形式的同伴回饋能夠幫助學生更準確地理解反饋信息，從而更有針對性地修改寫作，提高寫作品質。

2.2 人工智慧輔助同伴評價在寫作中的應用

人工智慧輔助同伴評價已被相關研究證明能夠顯著提高學生的回饋品質和寫作能力。與傳統的互評相比，人工智慧能更準確地識別寫作中存在的問題，並透過提供準確及個性化的回饋幫助學生有效改進寫作（Guo et al, 2024; Wang, 2020）。同時，這類工具還能有效增強學生的寫作信心與批判性思維，進一步促進他們的認知發展（孔維梁, 2023）。

3. 研究方法

3.1 研究對象

本研究共招募了一所中學的 27 名 13-14 歲的學生，均具備一定的英語寫作基礎。學生被隨機分配到實驗組（13 人）和對照組（14 人）。實驗組接受教師、傳統同伴和人工智慧同伴評分，而對照組只接受教師和傳統同伴評分。所有參與研究的學生均由同一位擁有豐富教學經驗的教師授課，以減少變量干擾。

3.2 研究流程

教學實驗為期 6 周，每週一次，每次 45 分鐘。

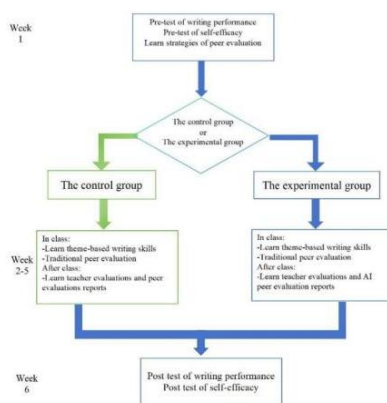


圖 1

3.3 研究工具

本研究的研究工具包括同伴評價表和自我效能感調查問卷。同伴評價表涵蓋內容評分（7 分）及語言評分（8 分），基於中考英語作文評分標準設計並經自身教師審核（見圖 2）。自我效能感問卷是基於 Shell（1989 年）開發的問卷，它由 15 個專案組成，採用李克特 5 點量表，問卷的 Cronbach alpha 值為 0.958，表明問卷是可靠的。而收集回來的相關實驗資料使用 SPSS 26.0 軟體進行分析。

同伴互評評分表	
內容 評分	1. 學校旅行的時間和地點 (共2分) 2分: 清楚提及時間和地點。 1分: 僅提及時間或地點中之一項。 0分: 未提及時間和地點。
	2. 旅行經歷 (共3分) 3分: 提及參觀農場並補充2項活動。 2分: 提及參觀農場並補充1項活動。 1分: 僅提及參觀農場, 未補充其他活動。 0分: 未提及旅行經歷或活動。
	3. 表達感受 (共2分) 2分: 提及至少2點感受。 1分: 提及1點感受。 0分: 未表達任何感受。
語言 評分	1. 邏輯 (1.5分) (1) 文章段落結構 (0.5分) 0.5分: 全文有明確的開頭、中間論述和結尾。且分為至少3段。 0分: 結構不完整, 無明顯分段。 (2) 段落主旨句 (0.5分) 0.5分: 每段包含清晰的主旨句, 概括段落大意。 0分: 缺乏主旨句或主旨句不明確。 (3) 過渡詞使用 (0.5分) 0.5分: 使用恰當的過渡詞連接句子 (如 First, Second, Besides, Finally 等)。 0分: 未使用過渡詞。

圖 2 同伴互評評分表 (1)

4. 研究結果

4.1 寫作能力差異分析

實驗組和對照組學生的寫作能力變化如表 1 和表 2 所示, 實驗組的前測成績與後測成績之間存在顯著差異, 寫作能力顯著提高。對照組學生的平均成績相對穩定, 寫作能力沒有顯著變化。在不同階段的比較中, 實驗組書寫水準呈上升趨勢。而對照組的書寫水準在實驗期間相對穩定。初步推斷人工智慧輔助同伴互評在提高寫作水準方面可能優於傳統的同伴互評方法。

表 1 對寫作能力的配對樣本 t 檢驗

組別	N	預試		事後測試		<i>t</i>	<i>p</i>
		平均值	SD	平均值	SD		
實驗組	13	9.97	3.71	13.01	1.03	-6.515	0.000**
對照組	14	8.44	3.76	8.74	4.08	-1.683	0.103

* $p < 0.05$ ** $p < 0.01$

表 2 不同階段寫作能力的配對樣本 t 檢驗

配對	組別	N	階段 1	階段 4	<i>t</i>	<i>p</i>
			平均值 (標清)	平均值 (標清)		
階段 1-階段 4	實驗組	13	10.00(1.86)	12.97(1.30)	-10.84	<0.001
	對照組	14	9.47(2.85)	10.22(2.57)	-4.13	<0.001

* $p < 0.05$ ** $p < 0.01$

4.2 自我效能分析

如表 3 所示, 對照組自我效能感前後測變化不明顯, 實驗組自我效能感顯著提高。根據獨立 t 檢驗結果, 如表 4 所示, 實驗組的後測自我效能感顯著高於對照組, 這進一步表明人工智慧輔助同伴評價對學生自我效能感的提升效果優於傳統同伴評價方式。

表 3 自我效能感的方差分析

組別	N	平均值	SD	<i>F</i>	<i>p</i>
實驗組	13	64	7.48	12.781	<0.001
對照組	14	46.35	11.47		

* $p < 0.05$ ** $p < 0.01$

表 4 自我效能感的獨立 t 檢驗

組別	N	平均值	SD	<i>t</i>	<i>p</i>
實驗組	13	64	7.48	7.301	0.000**
對照組	14	46.35	11.47		

* $p < 0.05$ ** $p < 0.01$

5. 結論

本研究探討了人工智慧輔助同伴評價與傳統的同伴互評方法對學生寫作水準、自我效能感以及學生對互評的態度的影響。經過為期四週的教學實驗，研究結果顯示，實驗組學生的寫作水準顯著高於僅採用傳統同伴互評方法的對照組。人工智慧輔助同伴評價透過個性化的及時回饋，幫助學生更準確地發現並改正寫作中存在的問題（Guo et al., 2024）。而傳統的同伴評價則主要依賴於學生間的主觀判斷，雖具有一定價值，但在準確性與個性化方面有所欠缺。兩者結合能更全面地理解和運用寫作技巧來提高寫作水準（Banihashem et al., 2024）。

此外，實驗組學生在自我效能感方面也有顯著提升，而對照組未呈現明顯變化。人工智慧輔助同伴評價所提供的即時且個性化的回饋，幫助學生更清晰地了解自己的優勢與不足，從而顯著提升其自我效能感。

参考文献

- 昂娟, & 陈媛. (2024). 大学生理解、认同与采纳同伴评语的实证研究. *荆楚理工学院学报*, 39(3), 1-8.
- 孔维梁, 于晓利, 韩淑云 & 邓敏杰. (2023). 人工智能赋能的可信同伴互评模型构建与验证. *现代远程教育研究*(03), 93-101+112.
- Banihashem, S. K., Kerman, N. T., Noroozi, O., Moon, J., & Drachsler, H. (2024). Feedback sources in essay writing: peer-generated or AI-generated feedback? *International Journal of Educational Technology in Higher Education*, 21(1).
<https://doi.org/10.1186/s41239-024-00455-4>
- Chang, C. Y. (2016). Two decades of research in L2 peer review. *Journal of Writing Research*, 8(1), 81 - 117. <https://doi.org/10.17239/jowr-2016.08.01.03>
- Gao, R., Merzdorf, H. E., Anwar, S., Hipwell, M. C., & Srinivasa, A. R. (2024). Automatic assessment of text-based responses in post-secondary education: A systematic review. *Computers and Education Artificial Intelligence*, 6, 100206.
<https://doi.org/10.1016/j.caeai.2024.100206>
- Guo, K., Pan, M., Li, Y., & Lai, C. (2024). Effects of an AI-supported approach to peer feedback on university EFL students' feedback quality and writing ability. *The Internet and Higher Education*, 63, 100962. <https://doi.org/10.1016/j.iheduc.2024.100962>
- Hsia, L., Huang, I., & Hwang, G. (2015). A web - based peer - assessment approach to improving junior high school students' performance, self - efficacy and motivation in performing arts courses. *British Journal of Educational Technology*, 47(4), 618 - 632.
<https://doi.org/10.1111/bjet.12248>
- Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, 81(1), 91 - 100. <https://doi.org/10.1037/0022-0663.81.1.91>
- Tseng, S., & Tsai, C. (2010). Taiwan college students' self-efficacy and motivation of learning in online peer assessment environments. *The Internet and Higher Education*, 13(3), 164 - 169. <https://doi.org/10.1016/j.iheduc.2010.01.001>
- Wang, Z. (2020). Computer-assisted EFL writing and evaluations based on artificial intelligence: a case from a college reading and writing course. *Library Hi Tech*, 40(1), 80 - 97.
<https://doi.org/10.1108/lht-05-2020-0113>
- Zhao, Y., Zhang, F., Schunn, C. D., He, P., Li, D., & Zhao, Y. (2023). Feedback, feedback-on-feedback and re-feedback: effects of written dialogic peer feedback on English as a foreign language writing. *Assessment & Evaluation in Higher Education*, 49(4), 523 - 538. <https://doi.org/10.1080/02602938.2023.2278017>