

生成式人工智能反馈对 EFL 学习者写作质量的影响——基于 25 项实验与准实验的元分析

The Impact of Generative AI Feedback on Writing Quality Among EFL Learners: A Meta-Analysis of 25 Experiments and Quasi-Experiments

李兴华¹, 张帝^{2*}

^{1,2} 浙江师范大学教育学院

*zhangdi0909@126.com

【摘要】 随着生成式人工智能(GenAI)在英语二语(EFL)写作教学中的广泛应用, 探讨其反馈成效对教育实践具有重要意义。本研究采用元分析方法评估了 GAI 反馈对写作质量的影响。结果显示, GenAI 反馈具有显著的正向促进作用。调节效应分析发现, 中学学段的干预效果优于大学学段; 12 周及以上的长周期干预效果最为突出; 且双向交互和范文对比模式能提供更深度的写作支持。基于此, 本研究建议教学中应实施长周期的专项干预, 并积极重构“师—生—机”三元反馈闭环, 以发挥人机协同育人的最大效能。

【关键词】 生成式人工智能; EFL 写作; 写作质量; 元分析; 自动化反馈

Abstract: With the widespread application of Generative AI (GenAI) in EFL writing instruction, exploring its feedback effectiveness is crucial for educational practice. This study employed a meta-analysis to evaluate the impact of GenAI feedback on writing quality. The results indicate that GenAI feedback has a significant positive promoting effect. Moderator analysis revealed that: the intervention effect at the secondary school level was superior to that at the university level; long-term interventions of 12 weeks or more were most prominent; and two-way interaction and model text comparison modes provided deeper writing support. Accordingly, this study suggests implementing long-term specialized interventions and reconstructing the "teacher-student-machine" ternary feedback loop to maximize the efficacy of human-machine collaborative education.

Keywords: Generative AI, EFL writing, writing quality, meta-analysis, automated feedback

1. 引言

写作能力在英语作为第二语言(EFL)学习者的语言发展中占有举足轻重的地位, 需要学习者熟练掌握写作的词汇、语法、内容和结构等多个维度(Tsai et al., 2024)。熟练的写作能力不仅可以帮助学习者清楚地表达观点、提升语言准确性与逻辑严密性 (Song et al., 2023), 还有助于未来的职业发展(Yang et al., 2025)。在写作教学中, 及时有效的反馈对提升学生的技能至关重要, 然而传统的教师反馈存在耗时、工作量大和个性化不足等问题(Lee et al., 2021)。

近年来, 生成式人工智能(GenAI)被广泛应用于 EFL 写作教学中, 旨在为学习者提供及时、个性化的反馈。多数研究证实了 GenAI 在 EFL 写作领域中提升学生写作技能、激发动机以及缓解写作焦虑等方面的积极作用(Song, 2023; Derakhshan, 2025)。但也有研究指出 GenAI 反馈与教师反馈的效果无显著差异, 甚至可能产生负面影响(Niloy et al., 2024)。此外现有综述多侧重于广义的人工智能反馈, 缺乏针对 GenAI 对写作质量具体效果的量化分析(Kaliisa et al., 2025;

Liu et al., 2025)。基于此，本研究试图通过元分析方法整合实验数据，探究生成式人工智能反馈对 EFL 学习者写作质量的影响，以为教育实践提供依据。具体问题如下：(1)生成式人工智能反馈对 EFL 学习者的写作质量影响如何？(2)哪些因素会显著影响生成式人工智能在 EFL 写作中的应用效果？

2.研究设计

2.1. 研究方法 with 工具

研究采用元分析方法，对国内外关于 GAI 反馈对 EFL 学习写作质量影响的相关文献进行统计分析，主要过程包括文献检索与编码、数据分析和证据解释。研究使用(CMA-v3.0)作为分析工具，并选取标准化均差(SMD)作为效应值。

2.2. 文献检索与筛选

本研究对发表于 CNKI 数据库、Web of Science 核心数据库、Scopus 数据库、ProQuest 数据库、IEEE 数据库和 Science direct 数据库中的文献进行检索。中文检索主题为“生成式人工智能”并含“英语写作”并含“反馈”，共检索到中文文献 35 篇。英文检索主题为“Generative AI”或“GenAI”或“Large Language Model”或“LLM”或“ChatGPT”或“Machine Learning”或“AWE”或“Automated Writing Evaluation”并含“feedback”并含“writing”并含“EFL”或“ESL”或“L2”或“second language”，发表年份和文献来源不限，共检索到 506 篇文献。

本研究设定的文献纳入标准为：(1)研究主题必须是生成式人工智能反馈对 EFL 学习者写作质量的影响(2)研究类型为实验研究或准实验研究(3)研究中必须包括实验组和对照组；(4)研究中提供了足够的数据信息可供提取和计算。具体检索流程如图 1 所示：

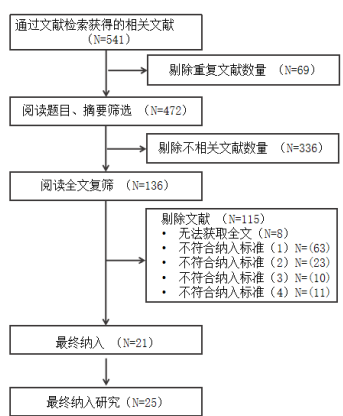


图 1 文献检索流程图

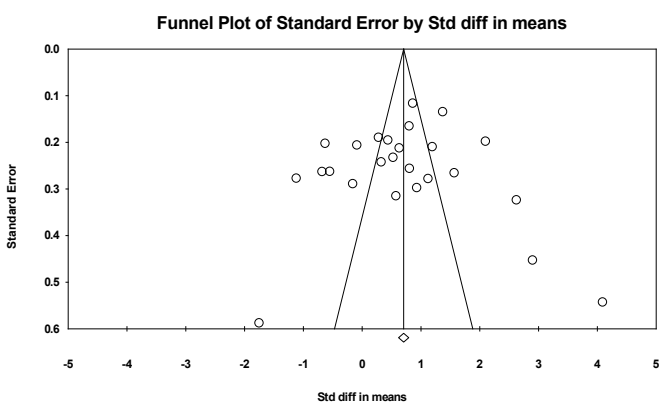


图 2 漏斗图

2.3. 文献编码

本研究对所筛选的 21 篇文献进行编码，编码信息包括文章基本信息、教育阶段、干预时长、反馈重点以及交互类型。其中，对反馈重点的分类参考了 Shute (2008) 关于形成性反馈的分类逻辑，将其划分为纠错、引导与综合三个维度；交互类型的分类则借鉴了 Steen-Ut heim 和 Wittek (2017) 的对话式反馈框架，将其划分为单向、双向以及范文对比模式。

3.研究结果

3.1. 发表偏倚和异质性检验

本研究通过漏斗图、失安全系数以及 Egger 线性回归检验发表偏差。从漏斗图的分布情况来看,大部分效应量较均匀地分布在对称轴中部两侧,基本呈左右对称(如图 2 所示)。Egger’s 回归分析结果显示，95%置信区间为[-5.52, 4.48], $p = 0.83 > 0.05$ ，表明不存在系统性的发表

偏倚。进一步从失安全系数来看,需 138 个额外的研究来抵消本元分析中发现的总体效应大小,远远大于 5k+10。综上,本研究样本间不存在明显的发表偏倚,结果比较稳定。异质性检验结果显示, $Q=380.55(p<0.001)$, $I^2=93.693$,说明本研究异质性水平较高。因此,本研究进一步开展调节效应分析,以识别导致异质性的显著影响因素,并采用随机效应模型进行后续分析。

3.2. 主效应检验

元分析结果表明, GAI 反馈对 EFL 写作质量具有显著的正向促进作用。合并效应量 $SMD=0.712$, 具有中等偏上的影响, 95%置信区间为 $[0.356, 1.067]$, $p<0.001$, 具有统计学显著性 (如表 2 所示)。

表 2 总体效应值表

模型	效应值和 95%的置信区间				双尾检验		异质性检验				Tau 平方		
	样本	点估计值	下限	上限	Z 值	P 值	Q 值	Df 值	P 值	I ²	Tau 平方	标准误	Tau 平方
固定效应	25	0.706	0.619	0.792	15.940	<0.001	380.546	24	<0.001	93.693	0.745	0.276	0.863
随机效应	25	0.712	0.356	1.067	3.920	<0.001							

表 3 实验情境的调节效应分析表

调节变量	维度	效应数	SMD	标准误	方差	95%置信区间		双尾检验		异质性		
						下限	上限	Z	P	Q	d _f	P
学段	中学	8	1.246	0.329	0.108	0.601	1.891	3.786	<0.001	4.001	1	0.045
	大学	17	0.456	0.218	0.048	0.028	0.884	2.088	0.037			
干预时长	1-4 周	9	0.694	0.353	0.125	0.001	1.387	1.964	0.050	7.916	2	0.019
	5-12 周	12	0.384	0.244	0.059	-0.094	0.861	1.575	0.115			
	>12 周	4	1.758	0.423	0.179	0.928	2.588	4.151	<0.001			
反馈重点	纠错	4	1.678	0.430	0.184	0.836	2.520	3.907	<0.001	9.125	2	0.010
	引导	5	1.266	0.644	0.414	0.004	2.527	1.967	0.049			
	综合	16	0.341	0.195	0.038	-0.042	0.723	1.747	0.081			
交互类型	单向	15	0.327	0.217	0.047	-0.098	0.753	1.509	0.131	5.909	2	0.052
	双向	8	1.296	0.359	0.129	0.593	1.999	3.613	<0.001			
	范文对比	2	1.468	1.150	1.322	-0.785	3.722	1.277	0.202			

3.3. 调节效应检验

本研究根据学习阶段、干预时长、反馈重点以及交互类型进行了调节效应分析，分析结果如表 3 所示。

3.3.1. 不同学习阶段对写作质量的调节作用

在学段方面,组间异质性检验结果显示 $Q = 4.001(p < 0.05)$,中学阶段的效应($SMD = 1.246$)显著高于大学阶段($SMD = 0.456$),表明较低学段的学生对智能代理反馈的依赖程度更高，且从中获得的效益更为显著。

3.3.2. 不同干预时长对写作质量的调节作用

在干预时长上，调节作用显著($Q = 7.916, p = 0.019$)。12 周及以上的长周期干预效果($SMD = 1.758$)显著高于短周期干预效果，这表明 GenAI 的教育价值并非即时显现，而是随着干预周期的延长和人机交互的深入而逐渐增强。

3.3.3. 不同反馈重点对写作质量的的调节作用

反馈重点具有极显著的调节作用($Q = 9.125, p = 0.010$)。结果显示，专注于纠错或引导的单一维度反馈效果均显著优于综合反馈。这从认知负荷理论的角度解释了反馈的有效性：针对性强的单一维度反馈能够降低学习者的认知负荷，使其能集中注意力解决特定维度的写作问题；而信息量过大的综合反馈可能导致认知超载，削弱了反馈的实际影响。

3.3.4. 不同交互类型对写作质量的调节作用

在交互类型上, 调节效应接近显著水平($Q = 5.909, p = 0.052$)。双向交互(SMD = 1.296)和范文对比(SMD = 1.468)的效应值明显高于单向反馈(SMD = 0.327), 说明通过多轮对话或范文对比, 学生能获得更有深度、更多元化的写作支持, 从而有效提升写作质量。

4. 结论与建议

本研究通过对 25 项研究的量化整合, 系统评估了生成式人工智能反馈对 EFL 学习者写作质量的影响, 结果表明, GenAI 反馈对提升英语写作质量具有显著的正向促进作用。调节效应分析显示, 中学学段的效应量显著高于大学学段, 表明较低学段的学习者对智能工具的依赖度更高, 受益也更明显。在反馈重点上, 相较于包含大量信息的综合反馈, 专注于纠错或引导的单一维度反馈由于降低了认知负荷, 展现出更优的干预效果。此外, 干预周期在 12 周以上的长期应用效果最为突出, 且双向交互和范文对比模式能为学生提供比单向反馈更有深度的写作支持。

基于上述研究结论, 本研究为提升 GenAI 在写作教学中的应用效能提出以下建议: 第一, 教学设计应遵循语言能力发展的渐进规律, 实施长周期的专项干预, 建议将 GenAI 反馈深度整合入学期课程体系中, 初期聚焦字词句层面的精准性, 中期引入思维结构引导和元认知提示, 从而稳步提升学生的学习效果。第二, 应积极重构“师—生—机”三元反馈闭环, 充分发挥 GenAI 在处理重复性纠错任务中的潜力, 使教师能将更多精力投入到深度点评、情感支持及写作引导上, 构建优势互补的协同反馈体系。

参考文献

- Derakhshan, A. (2025). Tracking the Effects of Gemini as a GenAI Tool on L2 Learners' Writing Proficiency and Anxiety: Latent Growth Curve Modeling Approach. *The Asia-Pacific Education Researcher*.
- Kaliisa, R., Misiejuk, K., López-Pernas, S., & Saqr, M. (2025). How does artificial intelligence compare to human feedback? A meta-analysis of performance, feedback perception, and learning dispositions. *Educational Psychology*, 0(0), 1~32.
- Lee, H.-S., Gweon, G.-H., Lord, T., Paessel, N., Pallant, A., & Pryputniewicz, S. (2021). Machine Learning-Enabled Automated Feedback: Supporting Students' Revision of Scientific Arguments Based on Data Drawn from Simulation. *Journal of Science Education and Technology*, 30(2), 168~192.
- Liu, Y., Kong, W., & Merve, K. (2025). ChatGPT applications in academic writing: A review of potential, limitations, and ethical challenges. *Arquivos Brasileiros de Oftalmologia*, 88, e2024.
- Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2024). Is Chatgpt a menace for creative writing ability? An experiment. *Journal of Computer Assisted Learning*, 40(2), 919~930.
- Song, C., & Song, Y. (2023). Enhancing academic writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology*, 14.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153~189.
- Steen-Utheim, A. T., & Wittek, A. L. (2017). Dialogic feedback and potentialities for student learning. *Learning, Culture and Social Interaction*, 15, 18~30.
- Tsai, C.-Y., Lin, Y.-T., & Brown, I. K. (2024). Impacts of ChatGPT-assisted writing for EFL English majors: Feasibility and challenges. *Education and Information Technologies*, 29(17), 22427~22445.

Yang, Y., Huang, L., Lin, W., Li, Y., Xu, Y., & Cheng, L. (2025). Enhancing Sustainable English Writing Instruction Through a Generative AI-Based Virtual Teacher Within a Co-Regulated Learning Framework. *Sustainability*, 17(19), 8770.