

Genio: 嵌入式编程教育中多模态 AI 智能体的设计与实验研究

Genio: Design and Randomized Controlled Trial of a Multi-modal AI Agent in Embedded Programming Education

王成梁^{1, 2}, 王浩名^{3, 2}, 李园园⁴ 刘星畅⁵, 陈小娇⁶, 林宇鹏^{7, *}

¹Australian Catholic University, Institute for Positive Psychology and Education

²华东师范大学, 教育学部, 教育信息技术学系

³清华大学, 教育学院

⁴上海师范大学, 分析测试与超算中心

⁵江苏师范大学, 人工智能与计算机学院 (智慧教育学院)

⁶浙江工业大学, 教育学院

⁷清华大学, 外国语言文学系

* linyupeng_research@126.com

【摘要】随着生成式人工智能的发展, 编程教育正由代码技能训练转向高阶思维培养。针对嵌入式编程教学中软硬件耦合度高、排错复杂及学习挫败感强等问题, 本研究基于人本主义理念与检索增强生成技术, 设计并开发了多模态智能教学助手 Genio。该系统集成 C++ 编程知识库、错误类目库与 Arduino 元器件视觉库, 通过问题分类与模块化并行流程, 实现情感支持、知识检索、引导式提示、困难诊断与视觉识别等功能。研究采用随机对照实验, 将 45 名被试分为 Genio 组与通用 LLM 组, 开展三轮实践干预。结果显示, Genio 组在编程知识水平与高复杂度任务的实践迁移能力上显著优于对照组; 在高阶思维方面, Genio 对算法思维、代码知识与创造性思维具有显著促进作用, 但对评估能力与批判性思维的提升尚未呈现显著优势。研究为多模态教学智能体设计及 AI 赋能编程教育提供了实证参考。

【关键词】人工智能智能体; 嵌入式编程; 编程学习; 随机对照试验; 高阶思维

Abstract

With the development of generative artificial intelligence, programming education is shifting from basic coding skill training toward the cultivation of higher-order thinking. To address the challenges of embedded programming education—such as strong hardware–software coupling, complex debugging processes, and learners’ heightened frustration—this study, grounded in humanistic learning theory and retrieval-augmented generation, designed and developed a multimodal intelligent teaching assistant named Genio. The system integrates a C++ programming

knowledge base, an error taxonomy database, and an Arduino component visual library. Through a problem classification mechanism and modular parallel workflows, Genio provides functions including affective support, knowledge retrieval, guided prompting, difficulty diagnosis, and visual recognition. A randomized controlled experiment was conducted with 45 participants assigned to either a Genio group or a general-purpose LLM group across three rounds of practical interventions. The results indicate that the Genio group significantly outperformed the control group in programming knowledge acquisition and demonstrated stronger transfer ability when solving high-complexity tasks. Regarding higher-order thinking, Genio significantly enhanced learners' algorithmic thinking, code knowledge, and creative thinking, while no structural advantage was observed in the development of evaluative ability and critical thinking. This study provides empirical evidence for the design of multimodal educational agents and for the application of AI in programming education.

Keywords: AI-Agent; Embedded programming; Programming learning; Randomized controlled trial; Higher-order thinking

1. 前言

过去十年,编程教育的重要性显著提升,已由单纯的技能训练转变为21世纪教育体系中的关键组成部分(Chen et al., 2020; Jing et al., 2024)。多国将编程教育纳入国家教育政策,以培养具备创新能力与问题解决能力的新一代公民(Tellhed et al., 2022)。随着生成式人工智能(GenAI)的兴起,编程教育在培养高素质人才中的战略地位进一步凸显,UNESCO等国际组织相继发布教育发展指南,强调在GenAI时代通过编程教育促进计算思维、批判性思维等高阶能力的发展(EDUCAUSE, 2023; UNESCO, 2025; Wang et al., 2025)。由此,编程教育已从面向少数技术精英的专业训练,转向面向全体学习者的核心素养培养。

在这一背景下,如何借助前沿技术优化编程教学、提升学习成效并促进高阶思维发展,成为研究关注的核心议题。信息与通信技术(ICT)、Web技术及虚拟学习环境的应用,显著拓展了编程教学的支持方式,为个性化学习与复杂问题解决提供了条件(Chen et al., 2020; Reisoğlu et al., 2017; Vermette et al., 2017)。GenAI的出现进一步推动了这一转型。已有研究表明,大语言模型(LLMs)在提升编程问题求解效率、学习动机及部分高阶思维方面具有潜力,但也引发了认知外包与学习依赖等问题(Chen et al., 2024; Jing et al., 2024; Yang et al., 2025)。

在此基础上,AI Agent作为能够感知环境、进行决策并自主行动的智能系统,逐渐受到教育研究关注(Lin & Yu, 2024)。已有研究显示,AI Agent支持的协作编程学习在提升学习成绩、自我效能与学习兴趣方面具有优势(Wang et al., 2025)。然而,现有研究多聚焦一般编程情境,对于软硬件高度耦合的嵌入式编程以及多模态学习支持机制的探讨仍然有限,相关实证证据亦有待补充。

鉴于此,本研究设计并开发了多模态编程教学AI Agent——Genio。该系统基于人本主义理念与检索增强生成(RAG)技术,以Arduino嵌入式编程为实践场域,构建C++编程知识库、错误类目录及元器件视觉数据库,使智能体具备多模态认知与协同问题解决能力,从而支持学习者完成复杂嵌入式编程任务并促进高阶思维发展。研究围绕以下问题展开:

RQ1: 在嵌入式编程的多模态学习情境中,Genio的知识库体系与工作机制如何设计?

RQ2: 相较于通用LLM,Genio是否能够有效提升学习者的编程知识与实践能力?

RQ3: 相较于通用LLM,Genio是否能够有效提升学习者的高阶思维能力?

2. 智能体设计

为回应嵌入式编程教学中多模态交互需求高、错误定位困难及学习挫败感强等问题，并在技术支持的同时体现人本主义的情感关怀（Chen et al., 2025），本研究构建了多模态编程教学 AI Agent——Genio。系统设计围绕两条主线展开：一是支撑多模态认知与精准反馈的知识库体系建设，二是以学习者为中心的智能体运行机制设计。

在知识层面，Genio 构建了一个结构化、语义化、多模态融合的知识体系，包括 C++ 编程知识库、编程错误类目库以及 Arduino 元器件图像知识库。其中，C++ 编程知识库以语法结构与程序逻辑为主线，系统梳理语言结构、条件判断、循环控制、函数调用及指针操作等核心内容，主要参考《C++ Primer》等权威教材，并按照由浅入深的学习路径进行组织。编程错误类目库则来源于真实教学情境，研究者在大学 C++ 程序设计课程中系统收集并编码了 69 名学生在 12 次课程作业中的高频错误，将个体性问题抽象为可复用的错误诊断模式，从而增强知识库对实际学习情境的适配性。

Arduino 元器件图像知识库以视觉感知为核心，由高分辨率元器件图像及配套文字说明构成，涵盖模块名称、功能特性、引脚定义、技术参数与典型应用场景。该设计使 Genio 能够在学生上传电路照片或元器件图像时，自动识别硬件类型、比对连线与引脚信息，并生成针对性的诊断与指导建议。相较于纯文本知识库，图像知识库更贴近嵌入式实验教学的实际需求，显著提升了系统在硬件排错与教学交互中的直观性。

知识库在数据结构上采用统一的元结构设计，通过向量化处理与统一索引机制，实现文本与图像的跨模态映射与关联检索。文本类知识条目经过清洗与语义嵌入处理，以支持自然语言查询；图像类知识则通过视觉特征提取并融合描述性文本，形成多模态向量表征，使系统能够实现“图文互查”。例如，当学生描述“这是一个三脚 LED 模块”并上传图片时，Genio 可通过视觉匹配定位相应元器件条目，并联动 C++ 知识库调用控制代码模板，实现软硬件知识的协同支持。

在机制层面，Genio 基于 Dify 平台开发，采用模块化与并行处理的系统架构。系统首先通过问题分类机制对学生输入进行分析，判断其需求类型（如情感支持、知识查询、调试引导或硬件诊断），并据此激活相应功能模块并行处理。系统核心模块涵盖情感回应、知识检索、引导式提示、学习困难诊断、视觉识别与教学模型生成等功能，其设计强调在技术正确性之外，兼顾反馈的解释透明性与语言温度，以维护学习者的主体地位。

Genio 的整体流程采用“分类—处理—汇总”的三段式架构（见图 1）。当学生提出问题后，系统首先对输入进行多维分析并完成分类；随后，一个或多个功能模块并行运行，例如在学生同时表达技术困难与情绪挫败时，系统可同步提供情感回应与技术支持；最后，各模块输出经由智能汇总节点整合，生成连贯、个性化的最终反馈（见图 2）。这种并行与汇总机制避免了单一回应路径的局限，使系统能够更全面地响应复杂学习需求。

图 1.
Genio 整体架构流程图

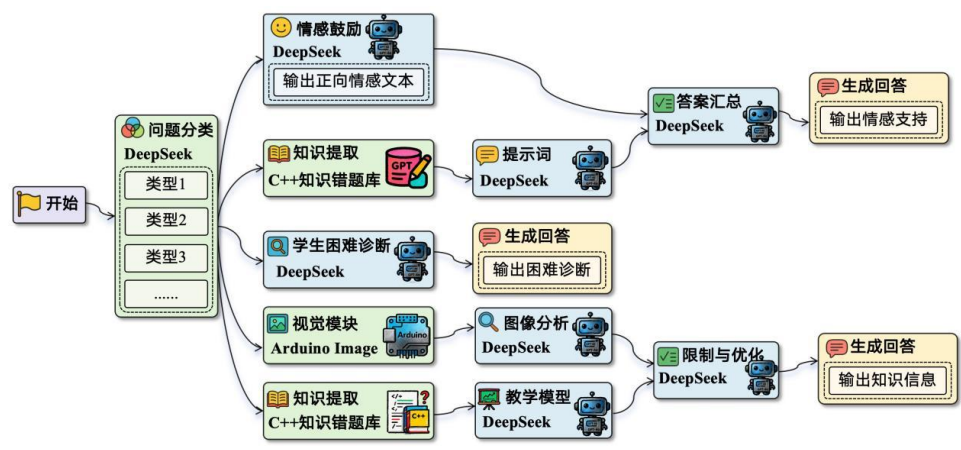
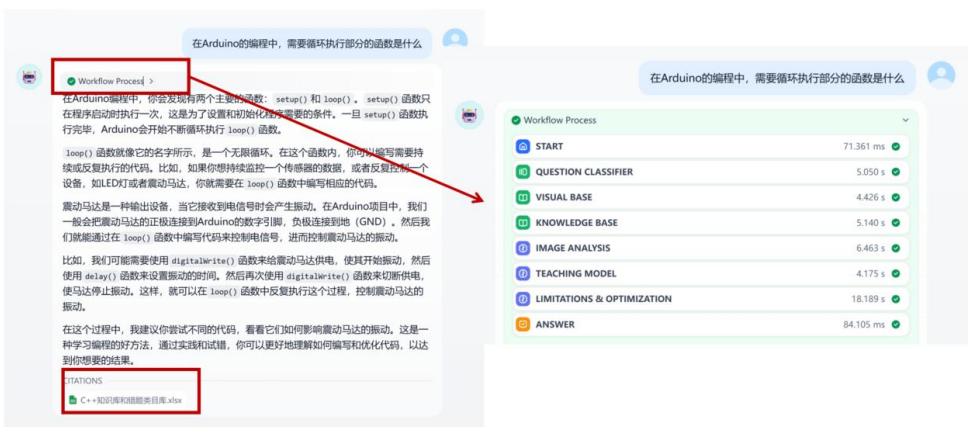


图 2.
完整工作流的示例



3. 实验设计与实施

3.1. 实验设计与被试招募

本研究通过设置实验组（Genio 组）与对照组（通用 LLM 组）的对比，考察 Genio 介入对学生编程学习成绩与高阶思维发展的影响，检验其在关键编程能力提升上的结构性优势。研究采用招募后随机分配的方式形成两组，并通过 C++知识测验与高阶思维前测验证基线一致性。正式实验以 Arduino 套件为载体，分三阶段完成楼梯 LED、智能风扇和迎宾防盗三项实践任务，配套统一的视频与操作指南。每阶段结束后实施等值知识测验，实践表现由两名评估者依据完成度评分，测试阶段禁用 LLM 与 Genio。实验结束后进行知识后测，并施测高阶思维与学习体验量表，通过统计分析评估教学成效。被试通过高校渠道招募，共 51 人参与，因中途退出最终获得有效样本 45 人（Genio 组 24 人，LLM 组 21 人），所有完整参与者均获得等值礼品奖励。

3.2. 实验材料

本研究设置三个连续教学干预阶段，分别组织三项基于 Arduino 套件的实践任务，并配套约 10–15 分钟的 MOOC 式教学视频作为学习脚手架。正式实验前，教师对 Arduino 套件的基本概念、常用元器件及编程控制原理进行简要介绍。随后，被试在观看教学视频后，依据

实践指南完成楼梯 LED 控制系统、智能电风扇交替控制系统及智能迎宾防盗系统三项任务，具体任务内容、涉及元器件及关键编程知识点。

在学习成效评估方面，研究共实施 4 次编程知识测试与 3 次实践成果测评。前测用于检验两组基线知识水平的一致性，其余三次测试在各阶段干预结束后进行，测试形式保持一致但内容依据教学材料设计。知识测试以选择题与填空题为主。实践成果根据任务完成度进行评分，被试在 Arduino 套件上完整实现目标功能即获得满分（100 分），中途出现偏差则按完成程度扣分。所有实践作品均由两名评估者独立评分并协商确认，以保证评分的客观性与一致性。

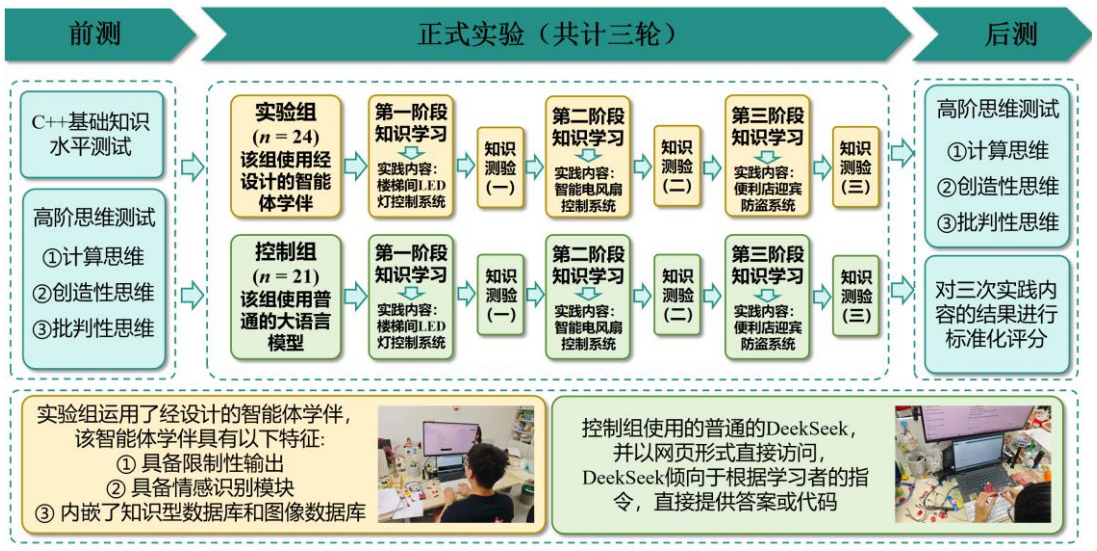
高阶思维采用计算思维、创造性思维与批判性思维三类量表进行测量，均为李克特 7 点量表，要求被试基于自身学习体验进行自评。计算思维量表参照 Kılıç et al. (2021)，从算法思维、代码知识与评估知识三个维度进行测量（Cronbach's $\alpha=0.76$ ）；创造性思维量表参考 Chiu and Hwang (2024)，包含 4 个题项（Cronbach's $\alpha=0.86$ ）；批判性思维量表参考 Lin et al. (2019)，同样包含 4 个题项（Cronbach's $\alpha=0.82$ ）。上述量表均在编程教育研究中具有良好的信度与适用性。

3.3. 实验流程

本研究的实验流程如图 3 所示，包含前测，正式实验和后测三个阶段。

图 3.

实验流程图



前测：在研究开始前，研究团队首先准备好了三轮实践活动的教学材料、实践指南与教学视频，C++基础知识水平测试与三轮正式测试试题，以及高阶思维评估和学习体验的量表，这些测试材料均是与一线教师多轮沟通后确定的，同时我们也进行了相应的预测试，这些操作确保了测试的难度是恰当的，以及提供给这些测试所需的时间是合适的。完成上述工作后进行实验前测，首先，先让学习者填写 C++基础知识水平测试，悉知被试在即将开展的学习活动前的先决知识水平，同时测量被试的高阶思维能力。

正式实验：正式实验是三轮实践活动，分别要求学生在学习相关知识后，利用 Arduino 组件完成楼梯间 LED 灯控制系统、智能电风扇交替控制系统和便利店智能迎宾防盗系统，完成

后将有评分员对其完成度进行评分。被试完成实践活动之后，将完成知识测验，在完成知识测验的过程中将无法使用 LLM 或 Genio，因为这是对学习者的知识能力掌握程度的评估。

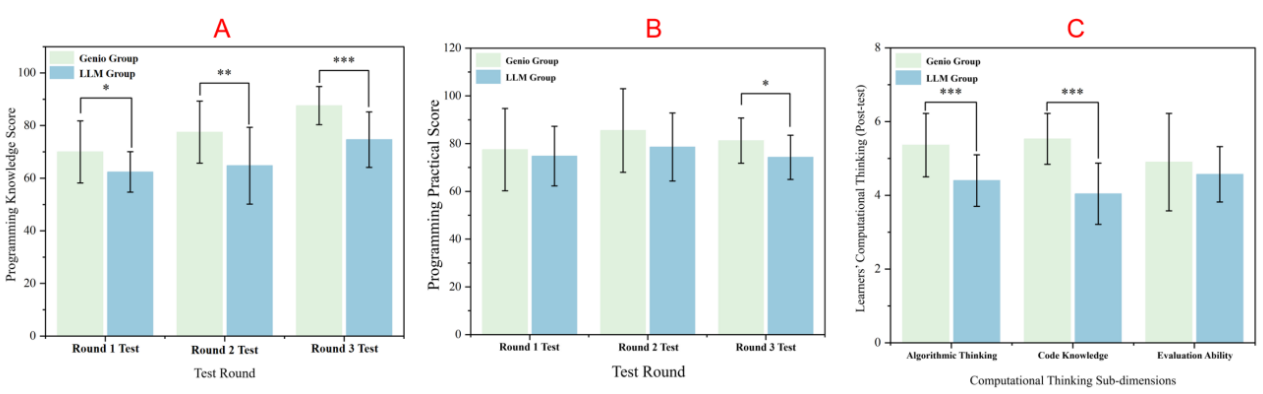
实验后测：完成三轮教学实践活动之后，被试将进行后测。完成作答后发放计算思维，创造性思维和批判性思维的量表。

3.4. 数据分析与结果

为检验两组被试的基线一致性，首先对实验前 C++编程知识水平进行独立样本 t 检验，结果显示 Genio 组与 LLM 组在前测中不存在显著差异 ($t = 0.65, p = 0.521$)，表明两组初始编程水平基本一致。

在此基础上，对三轮编程知识测试结果进行组间比较。结果表明，Genio 组在三轮测试中均显著优于 LLM 组，且差异随轮次递增（第一轮： $t = 2.53, p = 0.015$ ；第二轮： $t = 3.23, p = 0.002$ ；第三轮： $t = 4.72, p < 0.001$ ），具体结果见图 4A。这表明，相较通用 LLM，Genio 能持续提升学习者的编程知识掌握水平。

图 4.
学习者编程知识水平三轮测试对比



进一步对编程实践能力进行分析。结果显示，在前两轮低至中等难度实践任务中，两组未表现出显著差异（第一轮： $p = 0.551$ ；第二轮： $p = 0.418$ ）；但在第三轮高难度实践任务中，Genio 组的实践表现显著优于 LLM 组 ($t = 2.49, p = 0.017$)，相关结果见图 4B。这表明，Genio 的教学支持优势主要体现在复杂任务中的知识迁移与综合应用能力。

在高阶思维方面，研究对计算思维、创造性思维与批判性思维的后测结果进行了组间比较。计算思维分析结果显示，Genio 组在算法思维 ($t = 3.82, p < 0.001$) 和代码知识 ($t = 6.60, p < 0.001$) 两个子维度上显著高于 LLM 组，而在评估能力维度上未出现显著差异 ($t = 0.99, p = 0.327$)，具体见图 4C。

在创造性思维方面，Genio 组的后测得分显著高于 LLM 组 ($t = 3.24, p = 0.002$)，如表 1 所示。相较之下，两组在批判性思维后测中未表现出显著差异 ($t = -0.52, p = 0.608$)，如表 2 所示。

表 1.
学习者创造性思维后测组间比较

组别	N	均值±标准差	t 值	自由度(df)	p 值
Genio 组	24	5.03 ± 0.76	3.24	43	0.002**

LLM 组	21	4.32 ± 0.70
-------	----	-------------

表 2.

学习者批判性思维后测组间比较

组别	N	均值±标准差	t 值	自由度(df)	p 值
Genio 组	24	4.47 ± 1.23	-0.52	43	0.608
LLM 组	21	4.64 ± 0.99			

总体而言，与通用 LLM 相比，Genio 在提升学习者的编程知识、复杂实践能力、算法思维、代码知识及创造性思维方面表现出显著优势，但在评估能力与批判性思维的促进上尚未形成统计学上的显著差异。

4. 讨论与结论

本研究围绕嵌入式编程教学中知识理解困难、软硬件耦合复杂及高阶思维培养不足等问题，设计并实证检验了多模态教学 AI Agent——Genio，对其系统机制、教学成效及认知影响进行了综合分析。

针对 RQ1，研究验证了以多源知识库为认知底座的 AI Agent 设计在嵌入式编程教学中的有效性。Genio 将 C++编程概念、课堂高频错误类型与 Arduino 元器件特征整合为统一的可检索多模态知识体系，并通过检索增强生成实现错误定位、解释生成与情境化反馈的协同。其问题分类与模块并行机制按需调度情感支持、知识检索、提示引导与视觉识别，避免一次性给出完整答案，而以步骤化提示与解释性反馈引导学习者形成可检验假设，从而在降低挫败感的同时强化学习者的主体性与自我调节。

针对 RQ2，随机对照实验结果表明，Genio 在编程学习成效上显著优于通用 LLM。两组在前测中无显著差异，而在连续三轮干预后，Genio 组在编程知识水平上持续表现出显著优势，并在高复杂度实践任务中展现出更强的迁移与综合应用能力。这表明，系统化的知识引导与程序性支持未必立即转化为低阶操作优势，但在复杂问题解决中能够形成结构性差异，也揭示了单纯依赖 LLM 生成完整方案可能导致认知外包的风险。

针对 RQ3，Genio 对高阶思维的影响呈现差异化特征。其显著提升了算法思维、代码知识与创造性思维，主要得益于结构化知识支持与启发式交互机制；但对评估能力与批判性思维的促进未达显著，反映出当前以“问题—反馈”为核心的交互模式在支持反思性判断与元认知训练方面仍存在不足。由此，Genio 更适合作为促进程序性理解与知识迁移的“知识教练”，而非系统培养批判性判断的工具。

参考文献

- Chen, Y., Herskovitz, J., Matute, G., Wang, A. Y., Lee, S. Y., Lasecki, W. S., & Oney, S. (2020). EdCode: Towards personalized support at scale for remote assistance in CS education. In *Proceedings of the 2020 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 1-9). IEEE. <https://doi.org/10.1109/VL/HCC50065.2020.9127260>.

- Chen, X. J., Hu, Z. B., & Wang, C. L. (2024). Empowering education development through AIGC: A systematic literature review. *Education and Information Technologies*, 29, 17485-17537. <https://doi.org/10.1007/s10639-024-12549-7>
- Chen, X. J., Hu, Z. B., Li, Y. Y. & Wang, C. L. (2025). The journey of challenges and triumphs: a systematic literature review of the dynamic evolution of human-centered artificial intelligence in education. *Interactive Learning Environments*, 1-27. <https://doi.org/10.1080/10494820.2025.2472288>
- Chiu, M. C., & Hwang, G. J. (2024). Enhancing students' critical thinking and creative thinking: An integrated mind mapping and robot-based learning approach. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-024-12752-6>
- EDUCAUSE. (2023). 2023 EDUCAUSE Horizon Report: Teaching and Learning Edition. <https://library.educause.edu/resources/2023/5/2023-educause-horizon-report-teaching-and-learning-edition>
- Jing, Y. H., Dai, J., Wang, C. L., & Shadiev, R. (2024). Unleashing the power of virtual learning environment: Exploring the impact on learning outcomes through a meta-analysis. *Interactive Learning Environments*, 1-18. <https://doi.org/10.1080/10494820.2024.2335481>
- Jing, Y. H., Wang, H. M., Chen, X. J., & Wang, C. L. (2024). What factors will affect the effectiveness of using ChatGPT to solve programming problems? A quasi-experimental study. *Humanities & Social Sciences Communications*, 11, 319. <https://doi.org/10.1057/s41599-024-02751-w>
- Kılıç, S., Gökoğlu, S., & Öztürk, M. (2021). A valid and reliable scale for developing programming-oriented computational thinking. *Journal of Educational Computing Research*, 59(2), 257-286. <https://doi.org/10.1177/0735633120964402>
- Lin, H. C., Hwang, G. J., & Hsu, Y. D. (2019). Effects of ASQ-based flipped learning on nurse practitioner learners' nursing skills, learning achievement and learning perceptions. *Computers & Education*, 139, 207-221. <https://doi.org/10.1016/j.compedu.2019.05.014>
- Lin, Y. P., & Yu, Z. G. (2024). Systematically reviewing artificial intelligence-generated embodied virtual teachers: Effectiveness in diverse contexts and educational impacts. SSRN. <https://dx.doi.org/10.2139/ssrn.5703022>
- Reisoğlu, I., Topu, B., Yılmaz, R., Karakuş Yılmaz, T., & Göktaş, Y. (2017). 3D virtual learning environments in education: A meta-review. *Asia Pacific Education Review*, 18(1), 81–100. <https://doi.org/10.1007/s12564-016-9467-0>
- Tellhed, U., Björklund, F., & Strand, K. K. (2022). Sure I can code (but do I want to?). Why boys' and girls' programming beliefs differ and the effects of mandatory programming education. *Computers in Human Behavior*, 135, 107370. <https://doi.org/10.1016/j.chb.2022.107370>
- UNESCO. (2025). AI and the future of education: Disruptions, dilemmas and directions. <https://www.unesco.org/en/articles/ai-and-future-education-disruptions-dilemmas-and-directions>
- Vermette, L., Dembla, S., Wang, A. Y., McGrenere, J., & Chilana, P. K. (2017). Social CheatSheet: An interactive community-curated information overlay for web applications. *Proceedings of the ACM on Human-Computer Interaction*, 1(CSCW), 19. <https://doi.org/10.1145/3134737>
- Wang, C. L., Chen, Y. F., Hu, Z. B., Li, Y. Y., & Gu, X. Q. (2025). The journey of challenges and victories: Exploring the transformation action framework in the GenAI era from multifaceted policies. *Educational Technology Research and Development*, 1–43. <https://doi.org/10.1007/s11423-025-10535-5>

- Wang, H. M., Wang, C. L., Chen, Z., Liu, F., Bao, C. J., & Xu, X. L. (2025). Impact of AI agent-supported collaborative learning on the learning outcomes of university programming courses. *Education and Information Technologies*, 30, 17717–17749. <https://doi.org/10.1007/s10639-025-13487-8>
- Yang, T. C., Hsu, Y. C., & Wu, J. Y. (2025). The effectiveness of ChatGPT in assisting high school students in programming learning: Evidence from a quasi-experimental research. *Interactive Learning Environments*, 1-18. <https://doi.org/10.1080/10494820.2025.2450659>
- 