

# 生成式人工智能时代高校教育评价的证据危机、

## 过程确证与制度重构

黎银霞

(广州商学院, 广东 广州 511363)

**摘要:** 生成式人工智能正在深度嵌入高校课程学习、论文写作、科研训练与知识生产过程, 使传统以终结性成果为主要依据的教育评价面临证据效度危机。在人机共同生产条件下, 论文、代码、报告和项目方案往往由学生能力、模型能力、提示策略、平台资源和外部资料共同生成, 作品质量与学生真实能力之间的推断关系不再稳定。文章在高校数字化转型、人工智能教育应用、形成性评价和教育治理研究基础上, 按照研究前沿、问题提出、问题分析和解决方案的逻辑, 提出高校教育评价应从结果中心转向过程确证。过程确证并非以过程评价替代结果评价, 而是围绕明确的能力主张, 综合 AI 使用披露、生成过程记录、关键决策说明、现场表现和迁移性任务等多源证据, 对学生真实学习、实质贡献和责任承担进行交叉验证。高校应在任务分级、AI 使用披露、过程证据链、能力表现校验与责任数据治理之间建立制度闭环, 在创新、诚信、公平与育人价值之间形成新的评价平衡。

**关键词:** 生成式人工智能; 高校教育评价; 评价效度; 过程证据; 能力确证; 人机协同

DOI: doi.org/10.70693/202607148328

## Evidence Crisis, Process Confirmation, and Institutional Reconstruction of Higher Education Assessment in the Era of Generative Artificial Intelligence

Yinxia Li<sup>1</sup>

<sup>1</sup> School of Economics, Guangzhou College of Commerce, Guangdong, China.

**Abstract:** Generative artificial intelligence is becoming deeply embedded in university course learning, thesis writing, research training, and knowledge production, creating a crisis of evidential validity for traditional educational assessment that relies mainly on summative outcomes. Under conditions of human-AI co-production, papers, codes, reports, and project proposals are often jointly generated by students' abilities, model capabilities, prompting strategies, platform resources, and external materials. As a result, the inferential relationship between the quality of a submitted work and students' actual abilities is no longer stable. Drawing on research on university digital transformation, AI applications in education, formative assessment, and educational governance, this paper follows the logic of research frontiers, problem

**作者简介:** 黎银霞(1987—), 女, 博士, 研究方向为人工智能、高等教育

formulation, problem analysis, and solution design, and argues that higher education assessment should shift from outcome-centered evaluation to process confirmation. Process confirmation does not replace outcome assessment with process assessment. Rather, around clearly defined competence claims, it integrates multiple sources of evidence, including AI-use disclosure, records of the generation process, explanations of key decisions, on-site performance, and transfer tasks, in order to cross-validate students' authentic learning, substantive contribution, and responsibility-taking. Universities should establish an institutional closed loop among task classification, AI-use disclosure, process evidence chains, competence performance verification, and responsibility-data governance, thereby forming a new assessment balance among innovation, integrity, fairness, and educational value.

**Keywords:** generative artificial intelligence; higher education assessment; assessment validity; process evidence; competence confirmation; human-AI collaboration

## 一、前言

生成式人工智能对高校教育评价的影响,不能只从工具使用或学术诚信风险来理解。相关研究已经从高校数字化转型、人工智能教育应用、教学主体性、形成性评价和教育治理等方面形成了较为丰富的讨论,为重新思考高校评价制度提供了理论基础。

第一,高校数字化转型研究强调,人工智能并非单点工具更新,而是涉及教育理念、教学流程、治理机制、数据证据和伦理安全的系统性重构。杨宗凯等(2026)将高校数字化转型概括为底座、应用、以人为本、机制保障和数据证据等支柱,强调人工智能应深度嵌入教育教学全过程。Katsamakos等(2024)也指出,人工智能将同时作用于学习、科研、治理、学生服务和学术诚信等系统环节。Zawacki-Richter等(2019)对高校人工智能应用研究的系统综述表明,既有应用虽已覆盖教学支持、学习分析和智能辅导等领域,但对伦理、治理和评价影响的讨论仍需深化。

第二,生成式人工智能和推理语言模型研究揭示了学习活动本身的变化。Li等(2024)从基础能力、教育潜能和应用挑战等维度分析大语言模型适配教育的可能性,指出其在知识解释、问答反馈、写作辅助、编程支持和个性化学习中具有广泛应用空间。贾积有、陈昂轩(2026)进一步强调,推理语言模型能够通过可供检验的中间推理说明,引导学习者从答案获取转向问题分析、步骤检验与过程反思。由此可见,人工智能不仅改变学生交付什么成果,也改变成果如何形成。

第三,人工智能时代的教学理念研究提醒高校不能把技术应用简化为效率提升。张羽等(2026)提出人工智能撬动的系统性人本教学理

念,强调技术应服务人的发展、真实经验和学习共同体建设。王鉴、朱浩田(2026)从人机协同教学模式出发,讨论智能体介入后教师、学生、知识和工具等教学要素的重组。马香莲、陈正浩(2026)提出守拙教育哲学,提醒教育应保留慢思考、反复试错和发展性评价。赵世奇、王强(2026)从空间生产理论出发揭示数智教育空间中算法逻辑与教育公共性之间的张力。陆一(2026)则指出,AI降低研究生生产的时间成本后,更稀缺的是问题意识、理论直觉和责任感。

第四,形成性评价和教育治理研究为评价制度重构提供了基础。Scriven(1967)区分形成性评价与总结性评价,Bloom(1968)强调掌握学习中的反馈与矫正功能,Black和William(1998)系统论证课堂评价能够通过持续反馈促进学习。UNESCO(2023)关于生成式AI教育与研究的指导文件强调透明、伦理、责任和数据安全。Chan(2023)提出大学AI政策框架,Furze等(2024)提出AI Assessment Scale,为高校在不同任务中区分AI使用边界提供了操作化思路。Wang等(2024)对高校政策的分析表明,大学通常采取开放但谨慎的态度,强调透明使用、任务重构和学术诚信教育。

总体看,既有研究已充分揭示AI教育应用的能力基础、教学变革和治理风险,但仍存在三个可推进空间。其一,较多研究讨论AI如何提升教学效率,较少从评价理论角度分析AI如何改变成果与能力之间的证据关系。其二,较多研究关注学术诚信风险,较少提出能够同时识别真实学习贡献、规范AI使用并保护学生权利的评价制度。其三,较多研究从技术政策层面提出治理要求,较少深入处理学生主体性、教师专业判断、平台数据和责任归属之间的张力。

## 二、成果何以证明能力

生成式人工智能进入高校教育场域后，最先受到冲击的并不只是教学方法，而是教育评价赖以成立的证据关系。长期以来，课程论文、实验报告、代码作业、调研方案、项目展示和毕业论文等成果，通常被视为学生知识掌握、学习投入、能力水平与原创贡献的主要证据。教师依据最终作品的完整度、规范性、逻辑性和创新性作出判断，背后隐含着一个基本前提：作品主要由学生本人完成，因而作品质量能够较为可靠地反映学生能力。

生成式人工智能改变了这一前提。学生可以借助模型完成资料检索、概念解释、提纲设计、文本润色、代码生成、数据分析和论证框架搭建。最终作品越完整、语言越流畅、结构越规范，并不必然意味着学生理解越深入、投入越充分、贡献越真实。作品仍然重要，但作品不再自动构成能力的充分证据。高校教育评价由此面对一个更根本的问题：当成果由人机共同生成时，成果究竟如何证明能力？

这一问题并不意味着高校应把生成式人工智能排除在学习之外。人工智能已经成为新的学习基础设施和知识生产环境。真正需要回答的不是学生是否使用 AI，而是评价制度能否识别学生在使用 AI 过程中的问题意识、判断能力、修正能力、证据整合能力和责任承担能力。如果评价仍主要围绕最终成果展开，就可能把最容易被技术替代的形式完整度误认为能力，把工具产出误认为主体发展，把表层规范误认为真实学习。

因此，生成式人工智能对高校评价提出的核心挑战，不是简单的作弊检测问题，而是评价对象、评价证据和评价责任的系统性重构问题。高校需要在承认结果评价必要性的同时，把评价重点从单一作品判断扩展为对成果生成过程、工具使用过程、主体贡献过程和责任承担过程的综合识别。

## 三、证据危机与过程评价风险

传统高校评价能够以最终成果为主要依据，是因为它默认成果与学生能力之间存在相对稳定的关联。论文质量代表学术理解，代码质量代表编程能力，调研报告代表研究训练，项目方案代表综合应用能力。虽然这种推断过去也并不可靠，但在学生主要依靠自身知识、经验和劳动

完成成果的条件下，最终作品仍具有较高评价可用性。

人机共同生产使这一关系发生松动。成果不再主要由学生个人能力直接生成，而是学生、模型、提示词、平台资源、检索材料、同伴反馈和既有模板共同作用的结果。一个高质量成果可能来自学生的深度判断，也可能主要来自模型的语言整合；可能体现学生对问题的成熟把握，也可能只是模型对常见论述的顺滑组合。教师从成果质量反推学生能力时，会遭遇更强不确定性。

这种不确定性主要表现为三类失灵。第一是能力归因失灵。学生提交的成果可能达到较高形式水准，但教师难以判断其中哪些部分体现学生的概念理解、方法掌握和论证能力，哪些部分来自模型的通用生成能力。第二是贡献归属失灵。人机共制成果中，学生可能承担了选题、判断、修正和整合责任，也可能只是提交了模型生成结果；如果缺乏过程证据，评价难以区分实质贡献与表层操作。第三是责任配置失灵。AI 输出可能存在幻觉、偏见、伪造引文和逻辑漏洞，学生是否识别、修正并承担最终责任，应成为评价的重要内容；但传统结果评价往往只评判文本质量，忽视了工具使用中的判断责任。

由此产生的评价有效性危机，并不只是检测工具不够准确，而是评价证据结构发生了变化。若评价对象仍被限定为最终作品，就会把复杂的人机共同生产过程压缩为一个静态结果，进而遮蔽学生在过程中是否真正发生了学习、判断和成长。评价要重新有效，必须从作品文本扩展到成果形成过程，从成果评分扩展到能力确证。

但过程评价并非天然安全。首先是过程证据伪造风险。AI 不仅能够生成最终文本，也能够帮助学生补写过程说明、模拟反思日志、重构草稿演化甚至编造提示词记录。过程证据只有与课堂互动、口头答辩、现场任务和迁移性表现结合，才能提高确证能力。其次是认知不可见风险。学习过程中的理解、困惑、判断和顿悟并不总能被平台轨迹完整呈现。点击记录、版本历史和提示词摘要只能反映可记录行为，不能等同于真实认知过程。

再次是数据治理风险。过程确证需要适度收集草稿版本、学习日志、提示词摘要、平台轨迹和互动记录，但过度采集可能将教育评价变成持续监控。Chan (2023) 和 UNESCO (2023) 均强调，教育 AI 治理必须兼顾透明、隐私、安全、可解释和责任分配。过程证据越重要，数据

边界越需要明确。高校不能以评价改革之名无限扩大数据采集,更不能把学生置于持续被观察和被怀疑的状态。

此外,还存在公平分化和教师负担问题。不同学生在 AI 工具可及性、提示词能力、数字素养和学科资源方面存在差异。若评价制度简单鼓励高强度 AI 协作,可能使工具使用能力更强的学生获得优势;若简单禁止 AI,又可能忽视 AI 已成为学习和工作基础设施的现实。过程评价也意味着教师需要审阅更多证据、组织更多反馈和答辩,若缺乏制度支持,改革容易演变为额外劳动。

#### 四、高校教育评价的过程确证框架

面对生成式人工智能造成的证据危机,高校教育评价需要从结果中心转向过程确证。这里所说的过程确证,并不是以过程评价替代结果评价,也不是要求高校无限度记录学生学习轨迹,而是在人机共同生产条件下,围绕特定能力主张,综合 AI 使用披露、生成过程记录、关键决策说明、现场表现、迁移任务和反思日志等多源证据,对学生真实学习、实质贡献和责任承担进行交叉验证。

过程确证具有三个关键特征。第一,它以能力主张为中心,而不是以痕迹数量为中心。教师需要先明确某项任务评价什么能力:是基础知识掌握、独立写作能力、AI 协作能力、证据判断能力,还是研究设计能力。第二,它强调多源证据的相互印证,而不是迷信单一记录。提示词摘要、草稿版本、修改记录、课堂讨论、口头答辩和迁移任务都只能提供局部信息,必须组合使用。第三,它重视责任确认。学生可以合理使用 AI,但必须说明使用环节、判断依据和修改责任。评价不仅关注学生生成了什么,更关注学生为何如此生成、如何处理模型输出、是否能够为最终成果负责。

第一,建立任务分级制度。不同学习任务承担不同评价功能,不应采用同一 AI 使用规则。基础知识测验、现场写作和核心能力诊断,可限制或禁止 AI 使用,以确认学生基本掌握情况;课程论文、文献综述和项目报告,可允许 AI 辅助但要求披露使用范围;开放性研究、创新设计和跨学科项目,可鼓励人机协同,但必须强化过程证据与答辩环节。Furze 等 (2024) 提出的 AI Assessment Scale,为高校根据任务目标区分 AI 使用边界提供了有益参照。

第二,建立 AI 使用披露制度。学生在课程论文、毕业论文、项目报告和科研训练成果中,应说明 AI 参与环节、使用目的、采纳范围和修改责任。披露内容可包括资料检索、提纲生成、语言润色、代码调试、数据分析、观点生成等。披露不是为了简单惩罚,而是把 AI 使用纳入学术规范,使学生明确可以借助工具,但必须承担判断责任。

第三,建立过程性证据链。评价材料可包括选题说明、资料来源、提纲演化、草稿版本、提示词摘要、修改记录、课堂讨论、口头答辩和反思日志。高阶任务应要求学生解释关键选择:为什么这样提问,为什么采纳或拒绝模型输出,如何处理矛盾证据,最终观点如何形成。过程证据链的目的不是让学生提交更多材料,而是让教师能够围绕核心能力进行有针对性的判断。

第四,建立能力表现校验机制。过程证据仍可能被补写和伪造,因而必须通过现场表现进行校验。课程可设置简短口头答辩、随机追问、现场修改、迁移性小任务或案例解释,检验学生是否真正理解成果内容,是否能够说明关键概念、方法和判断依据。能力校验不是增加惩罚性审查,而是把评价从静态成果转向动态表现,使学生对自己提交的成果拥有可解释、可辩护和可迁移的掌握。

第五,建立责任与数据治理机制。学生承担披露、判断和最终提交责任;教师承担任务设计、评价解释和反馈责任;学校承担制度供给、教师培训、申诉处理和数据保护责任;平台应在隐私、安全和可解释方面接受规范。过程证据采集应遵循最小必要原则,只收集与评价目标直接相关的材料,并明确用途、保存期限、访问权限和删除机制。教育评价的数字化不能以牺牲学生隐私和师生信任为代价。

由此,高校教育评价的制度重构并非简单把结果评价替换为过程评价,而是在结果、过程、主体和责任之间建立新的关系。结果仍然重要,因为教育评价最终要判断学生是否形成能够外化和迁移的能力;过程也重要,因为 AI 时代的结果不能自动证明能力。制度设计的关键,在于用适度、相关、可解释、受约束的过程证据,支撑对学生能力和贡献的合理判断。

#### 五、结论与讨论

生成式人工智能并非简单破坏高校教育评价,而是迫使高校重新思考何为学习、何为能

力、何为原创贡献。传统结果评价的问题在于，它把最终作品视为能力的充分证据；AI时代的评价改革则要求高校看到作品背后的提问、判断、修正、反思和责任承担。过程确证的价值，不是记录更多学习痕迹，而是围绕能力主张建立多源、可验证且受权利约束的证据链。

高校教育评价的过程转向应坚持三项原则。其一，承认AI已经成为学习基础设施。评价制度不应回避学生使用AI的现实，而应通过任务分级和使用披露引导合理使用。其二，重视过程证据，但防止数据监控泛化。过程证据的价值在于支持能力确证，而不是把学生置于持续监控之下。其三，强化教师专业判断，同时提供制度、平台和量规支持。AI时代的评价改革不能以增加教师隐性劳动为代价，而应通过制度化支持使教师能够把评价重点放在学生真实成长上。

未来高校评价不应只问学生是否完成了一个漂亮成果，还应追问学生是否提出了有价值的问题，是否理解工具边界，是否能够判断证据质量，是否能够修正模型错误，是否能够在新的情境中迁移运用，并是否愿意为自己的学术表达承担责任。未来研究可继续开展课程层面的实证检验，比较不同AI使用披露方式、过程证据组合和答辩机制对学习质量的影响；也可围绕不同学科任务建立更细化的过程确证量规，形成兼顾教育价值、学术诚信、数据权利和技术创新的AI评价治理模型。

#### 参考文献

- [1] 杨宗凯, 程浩, 吴龙凯. 高校数字化转型的基本框架及实践审思[J]. 北京大学教育评论, 2026 (1): 51-65. DOI: 10.12088/pku1671-9468.202601004.
- [2] Katsamakos, E., Pavlov, O. V., & Saklad, R. (2024). Artificial intelligence and the transformation of higher education institutions. arXiv:2402.08143.
- [3] Zawacki-Richter, O., Marin, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16, 39.
- [4] Li, Q., Fu, L., Zhang, W., Chen, X., Yu, J., Xia, W., Zhang, W., Tang, R., & Yu, Y. (2024). Adapting large language models for education: Foundational capabilities, potentials, and challenges. arXiv:2401.08664.
- [5] 贾积有, 陈昂轩. 推理语言模型赋能教育的实践价值、风险挑战与发展进路[J]. 北京大学教育评论, 2026 (1): 66-77. DOI: 10.12088/pku1671-9468.202601005.
- [6] 张羽, 郝展欣, 覃菲. 人工智能撬动的系统性人本教学理念[J]. 北京大学教育评论, 2026 (1): 78-90. DOI: 10.12088/pku1671-9468.202601006.
- [7] 王鉴, 朱浩田. 人工智能时代人机协同教学模式的建构[J/OL]. 高等教育研究, 网络首发, 2026-05-27. <https://link.cnki.net/urlid/42.1024.G4.20260526.2221.008>.
- [8] 马香莲, 陈正浩. 迟钝的自由: 人工智能时代“守拙”教育哲学的价值意蕴[J]. 教育理论与实践, 2026 (16): 3-12.
- [9] 赵世奇, 王强. 数智时代教育空间的逻辑转向与生态重塑——基于空间生产理论的审视[J]. 教育理论与实践, 2026 (16): 20-26.
- [10] 陆一. AI时代的研究效率与评价转向[J]. 复旦教育论坛, 2026, 24 (2): 1. DOI: 10.13397/j.cnki.fef.2026.02.011.
- [11] Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum Evaluation*. Rand McNally.
- [12] Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1-12.
- [13] Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- [14] UNESCO. (2023). Guidance for generative AI in education and research. UNESCO.
- [15] Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. arXiv:2305.00280.
- [16] Perkins, M., Furze, L., Roe, J., MacVaugh, J. The AI Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment[J]. *Australasian Journal of Educational Technology*, 2024, 40(4): 1-19.
- [17] Wang, H., Dang, A., Wu, Z., & Mac, S. (2024). Generative AI in higher education: Seeing ChatGPT through universities' policies, resources, and guidelines. *Computers and Education: Artificial Intelligence*, 6, 1002.

