生成式人工智能辅助学习中的能力替代风险与评价机制重构

尹彦浩*

青岛工学院，山东 青岛 266300

摘 要：生成式人工智能能够改善学习任务的完成效率，但辅助条件下的高质量产出是否意味着学生获得了相应能力，仍须结合学习过程和评价证据判断。本文以认知负荷、认知卸载、自我调节学习及评价效度理论为基础，对相关实验研究、元分析和教育评价文献进行问题导向的概念综合。分析认为，能力替代风险形成于人工智能承担关键目标活动、学生缺少对反馈的自主加工、评价以辅助作品推断独立能力的结合情形。现有实验同时报告了人工智能辅导的学习收益与无约束使用的不利结果，说明工具设计、学习者基础与测量条件不可忽略。基于此，本文区分辅助任务表现、学习过程中的认知责任和独立迁移表现，提出与学习目标相匹配的分层证据评价思路，并讨论其实施条件与检验路径。研究为解释人工智能介入后学习成绩与能力形成的关系提供了分析框架。

关键词：生成式人工智能；学习评价；认知卸载；独立能力；评价效度

Risks of Cognitive Substitution and the Redesign of Learning Assessment in Generative AI-Assisted Learning

YIN Yanhao*

Qingdao Institute of Technology, Qingdao, Shandong 650500

Abstract: Generative AI can improve the efficiency and quality of students' task completion, but successful AI-assisted performance does not necessarily demonstrate independently acquired competence. Drawing on cognitive load theory, cognitive offloading, self-regulated learning, and assessment validity, this conceptual paper synthesizes findings from experimental studies, meta-analyses, and educational assessment research. It argues that the risk of cognitive substitution arises when AI performs cognitive operations central to the intended learning outcomes, learners do not meaningfully process the feedback, and assessment practices infer independent competence from AI-assisted products. Evidence of both beneficial AI tutoring and adverse effects of unrestricted use indicates that tool design, prior knowledge, and outcome measures matter. The paper distinguishes AI-assisted task performance, learners' cognitive responsibility during the process, and independent transfer performance, and develops an assessment rationale that aligns these sources of evidence with specific learning goals. The framework links learning goals, task conditions, and assessment evidence to guide the design and examination of assessment practices in AI-mediated educational settings.

Keywords: generative artificial intelligence; learning assessment; cognitive offloading; independent competence; assessment validity

DOI：https://doi.org/10.65192/fe9jq827

© 2026 作者

本文依据 CC BY 4.0 许可协议开放获取

1 引言

生成式人工智能进入学习场景后，学生可以在提出问题的同时获得完整解释、示范解答和修改建议。与传统检索工具不同，这类系统不仅提供资料，还可能直接承担分析、论证和表达等学习任务中的关键步骤。Yan 等指出，生成式人工智能扩大了个性化支持和即时反馈的可能性，也改变了教师判断学习结果的前提条件[1]。当学生提交的作品已经包含工具的认知贡献时，作品质量与学生实际掌握的能力之间不再具有理所当然的对应关系。

作者简介：尹彦浩（1996—），男，硕士研究生，助教，主要研究方向为外国语言文学，外语教育
通讯作者：尹彦浩，yinyanhao1996@163.com

现有研究呈现了不同教学配置下的学习收益与风险。Han 等对实验及准实验研究开展元分析，报告生成式人工智能干预总体上具有积极学习效果，但不同研究之间存在显著异质性[2]。Kestin 等在大学物理课程开展的随机对照试验发现，采用教学法设计的人工智能辅导能够在特定条件下取得优于课堂主动学习安排的测验表现[3]。与之形成对照的是，Bastani 等在高中数学场景中发现，无教学保护的辅助工具改善了练习成绩，却使撤除工具后的独立考试成绩下降；设置引导性提示后，这种不利影响明显减轻[4]。这些结果提出了一个共同问题：相似技术为何在不同教学配置下对应不同的能力结果。

这种差异同时提出了学习评价问题。在传统书面作业中，教师通常依据学生作品推断其知识理解与分析能力；当生成、组织和修订可以由外部系统完成时，作品所体现的能力来源变得混合。即使某种评价能够准确判断文本质量，也未必能够回答学生是否独立完成了核心推理。已有元分析发现，生成式人工智能对学习成绩和语言学习等指标有促进作用，但对元认知结果的作用并不一致[5]。若不说明测量对象，任务分数、能力发展和人机协作熟练程度很容易被混为一谈。

本文以认知加工和评价效度的关系为切入点，讨论两个相互联系的问题：生成式人工智能在什么条件下可能替代学习者应当完成的认知活动；教学评价又应如何识别工具辅助表现与学生能力之间的区别。研究采用问题导向的理论分析，将经典学习理论与已经发表的实验研究、元分析及评价研究加以比较，重点分析不同研究所测结果的差异，从而提出可供教学设计和后续实证研究检验的解释框架。

2 理论基础

2.1 认知负荷与认知卸载

从学习心理学看，降低完成任务所需的认知负担不一定会降低学习质量。认知负荷理论强调，学习者的加工资源有限，某些问题解决活动可能占用本应用于知识结构形成的资源[6]。当工具承担繁复检索、基础排版或重复性运算时，学生有可能将注意力转向概念比较和证据解释。这一作用取决于自动化活动与课程目标的关系。对于学习数据分析的初学者，选择统计模型、解释参数或发现异常可能正是学习目标；若这些步骤由系统直接给出，任务效率的提升便不能直接说明学习效果。

认知卸载研究将个体借助外部环境减少内部加工需求的行为视为普遍认知策略，并指出卸载选择还受到学习者对自身能力判断的影响[7]。这一理论能够解释学生为何倾向于将耗费精力的步骤交给人工智能。决定其教育意义的是卸载发生在何处、学生是否保留必要的理解活动，以及是否仍有独立运用知识的机会。依赖程度的评价需要识别被替代的具体活动，并考察这种替代与能力形成的关系。

2.2 认知投入与自主调节

学习活动还存在可见操作与实际认知加工的差别。Chi 和 Wylie 提出的 ICAP 框架根据学习者行为区分被动、主动、建构和互动等投入方式，为分析学习活动的深度提供了依据[8]。在人工智能交互中，学生反复发送提示语并不必然意味着深入学习；如果互动只是在要求系统重新表述标准答案，学生可能几乎没有形成新的解释。相反，当学生提出自己的假设，针对系统反馈修改论据，并能说明修改理由时，交互才可能与建构性学习活动相联系。认知投入的判断应结合具体任务、学生的解释及其回应质量。

自我调节学习理论进一步关注学习者如何设定目标、监控理解并调整策略。Panadero 对不同自我调节学习模型的综述说明，监控与反思不是学习结束后的附加步骤，而是连接策略使用与学习结果的重要过程[9]。人工智能能够提供即时答案，却不能自动保证学生识别自己是否理解。若学生将答案的流畅程度当作正确程度，可能停止检查前提、证据和推理；如果教师要求学生先尝试、再比较并解释不同解法，工具反馈则更有机会进入自我调节循环。由此，工具使用顺序本身就可能改变学生承担的认知责任。

2.3 学习支持与能力形成

教学支持的价值并不在于替学生持续完成任务，而在于帮助其逐步完成原先难以独立完成的活动。Wood 等关于辅导的经典研究说明，支持需要结合学习者当前能力调整，逐步使问题解决回归学习者自身[10]。生成式人工智能提供了低成本、可重复的支持渠道，但其响应机制并不天然具有教学递进性：直接给出完整答案可能满足即时需求，却跳过学习者最需要练习的步骤。教师制定的提示规则、错误反馈方式和撤除支持的时机，因而成为教学设计的一部分。

对“学会了什么”的判断还需要区分当场成功和延续性能力。本文将辅助任务表现界定为学生在可以使用人工智能的条件下完成任务的质量；将独立能力界定为学生在不依赖该工具完成目标所要求的关键认知活动的能力；将人机协作能力界定为学生能够合理分配任务、核验结果并对最终决定负责的能力。三者可能相关，却不属于同一个评价对象。独立能力的评价着重考察目标认知活动由学生完成的程度，人机协作能力的评价则关注学生如何运用基础知识核验并使用工具结果。评价时应先明确课程培养哪一种能力，再选择相应的表现证据。

3 替代风险

3.1 关键认知活动的外移

能力替代风险首先表现为任务完成过程与目标认知活动发生分离。以论证写作为例，如果教学目标是训练学生建立论点与证据之间的关系，那么由人工智能直接生成完整论证，便可能使学生在没有实际组织证据的情况下交出结构完整的文章。外部文本质量可以提高，学生需要练习的推理却未必发生。若教学目标改为批判性评价机器生成的论证，同样的生成行为又可以成为学习材料。两种情境的区别在于模型执行的步骤与学习目标是否重合。

Bastani 等的研究为这一判断提供了有条件的实证支持：学生使用无教学保护的 GPT 工具时练习成绩改善，但在随后的无工具考试中表现下降；采用教师设计提示的辅导方式则缓解了这种差异[4]。该研究表明，在所考察的高中数学任务中，直接提供答案与提供教学支架对应不同的独立学习表现。人工智能系统在任务中的位置，比抽象的“是否使用人工智能”更接近影响学习的可操作变量。

另一方面，Kestin 等在大学物理教学中发现，依据教学原则构建的人工智能辅导可以促进特定测验中的知识掌握[3]。两项研究分别呈现了无约束答案供给的风险与经过教学设计的对话支持的潜力。学习对象、基线能力、教师参与、对照条件与测验安排均可能影响结果。比较这些结果时，应将工具安排、教学过程和测验条件相互联系。

3.2 反馈使用中的判断让渡

反馈及时并不等于反馈有效。Shute 将形成性反馈的作用指向学习者思维或行为的改变，并指出反馈类型、呈现时机和学习者特征会影响效果[11]；Hattie 与 Timperley 也强调，反馈能否支持学习取决于它回答什么问题以及如何被使用[12]。人工智能可以随时给出修改方案，但如果学生将“接受建议”误当作“理解建议”，反馈只是在作品层面发生了作用。判断让渡可以从学生能否说明建议成立的条件、检验信息来源并识别不适用的情况加以观察。

例如，人工智能可能建议修改一份研究报告的统计分析。学生若只替换结果段落，作品改善主要体现于结果描述；如果学生能够检查变量类型、解释方法选择，并指出建议中的限制，工具反馈才更可能促进学科判断能力。教师可以选取关键修改要求学生解释，将修改理由和证据核验作为认知投入的观察依据。

Barcaui 在一项针对大学生的随机对照研究中报告，使用 ChatGPT 辅助学习的组别在 45 天后的知识保持测验中成绩低于传统学习组[13]。这一结果对应特定课程内容与使用安排下的延迟保持，提示即时任务表现与知识保持需要在不同时间点观察。若评价只覆盖工具使用期间的即时产出，就可能遗漏这种差异。

3.3 风险形成的条件

能力替代风险受到先备知识的制约。具有领域知识的学生更可能识别工具回答中的概念漏洞、遗漏条件与错误引用；初学者则可能因缺乏判断依据而高估答案可信度。相同的建议可能对前者构成高效的比较材料，对后者却成为无法检验的现成结论。因此，课堂规则不能只规定“允许”或“禁止”，还应考虑学生当前是否具备核验所需的基础知识，以及教师是否提供了必要的验证资源。

任务类型同样重要。在学习早期，概念形成、推理练习和诊断性测验通常需要保留学生的独立加工；到了综合应用阶段，适度使用专业工具可以成为真实能力的一部分。Deng 等对实验研究的元分析发现，ChatGPT 干预可以改善部分学习结果，同时降低任务所需的心理努力[14]。降低努力既可能意味着低效负担被消除，也可能意味着关键思维活动被外包。仅凭自报的轻松程度或满意度，很难判断具体发生了哪一种变化。

教学支持方式进一步改变风险。先展示完整答案、在学生尝试后提供提示、仅针对错误环节给予反馈，这些安排可能引发不同的学习行为。只要教师没有区分哪些步骤属于目标能力，人工智能的可用性就可能改变练习任务的实际认知要求。因而，风险由任务目标、学习者基础、反馈结构和可用资源共同形成。

4 评价重构

4.1 从作品质量转向证据效度

生成式人工智能介入后，评价首先需要重新审视分数解释是否成立。Kane 提出，评价效度不仅关乎评分是否一致，更关乎依据分数所作的推断及其用途是否有充分证据支持[15]。教师可以用统一量规准确评价一篇报告的论证质量，但若无法区分学生与工具分别承担的认知活动，就不宜据此直接推断学生能够独立完成相同论证。这一推断需要补充学生独立组织证据和完成论证的表现。

国内研究已经注意到这种转变。沈书生、祝智庭讨论 ChatGPT 类产品对学习评价的影响，强调从答案与知识呈现转向思维过程和问题解决[16]。这一判断提示了改革方向，但评价机制仍须回答更具体的问题：教师需要收集什么证据、采用何种任务验证，以及在多大程度上可以解释学生的能力。过程材料也可能由工具加工，教师需要判断学生在关键选择和修订中承担的认知活动，以及这些活动与学习目标的关系。

过程证据与结果证据应依据评价用途相互配合。Black 与 Wiliam 关于形成性评价的研究说明，学习证据只有进入反馈和教学调整，才能发挥促进学习的作用[17]。因此，本文主张依据评价用途组织不同证据：面向学习改进，重在识别学生在哪一步遇到困难以及反馈如何起作用；面向课程达标，重在确认学生是否达到规定能力；面向人机协作，重在核验学生对工具结果的判断与责任。用途不同，所需证据和工具使用规则也应不同。

4.2 区分三类学习证据

第一类是辅助任务表现，回答“学生在规定工具条件下能把任务完成到什么程度”。它能够评价信息整合、协作决策和作品质量，对于强调专业工具应用的课程尤其重要。独立推理能力则需要学生在相应独立任务中的表现作为证据。第二类是过程中的认知责任，回答“哪些关键选择由学生作出，学生如何检查并修订工具的建议”。可以通过关键节点的草稿、证据核验说明和针对性口头解释收集。第三类是独立迁移表现，回答“在工具撤除或情境变化后，学生还能完成哪些核心活动”。三类证据各有解释范围，应围绕教学目标组合使用。

独立迁移评价需要在保留核心能力要求的同时改变应用情境。Barnett 和 Ceci 指出，迁移研究需要明确知识内容和应用情境之间的距离，不同迁移任务不能简单视为同一种结果[18]。若学生刚练习过某道题，再完成形式几乎相同的无工具题目，更多反映即时再现；若评价旨在判断迁移，则应在保持核心能力要求的前提下改变部分情境条件，并控制难度差异。根据课程目标，还可以设置延迟测验，以区别短暂记忆与较为稳定的掌握。

表 1 将学习目标、工具角色与评价证据相互联系。评价任务应为学生提供展示目标能力的机会，评分解释应对应证据产生的学习条件。

表 1 生成式人工智能辅助学习的目标、工具角色与评价证据

主要学习目标	人工智能的适宜角色	优先观察的评价证据	需要防范的推断偏差
基础概念与独立推理	针对错误提供提示，核心推理由学生完成	无工具解释、关键步骤重建、延迟任务	以辅助解答的正确率推断独立掌握
论证写作与资料核验	生成备选观点、提供反馈、辅助检索	证据与论点匹配、修订理由、错误识别	将语言流畅度等同于论证能力
人机协作与专业应用	在明确责任下参与分析、生成和修订	工具选择依据、结果复核、决策责任	把工具产出全部归因于学生，或否认协作能力

4.3 评价任务的组织方式

在单元教学中，可以通过同一主题的连续任务减少“作品与能力不相符”的判断困难。以大学生论证写作为例，教师首先提供一则包含多方观点的现实议题，要求学生独立写出立场、核心依据和可能的反例；随后允许人工智能帮助寻找论证漏洞、形成替代解释；最后学生提交修订文本，并说明一项被采纳和一项被拒绝的建议。过程记录用于考察学生如何判断、采纳和修订建议。

如果课程目标还包括独立论证，教师可以安排短时的无工具任务，要求学生面对相近但不相同的材料重新提出论点，并在必要时进行口头追问。相反，如果课程明确培养人机协作写作能力，则可以允许工具持续参与，但应评价来源核验、论证责任与错误纠正。评价条件必须在任务开始前告知学生，不能先鼓励使用工具，再用模糊的“人工原创程度”标准追溯扣分。

评分时应避免把所有证据机械合成为一个总分。辅助作品、过程判断与独立表现可能服务于不同目标，先给三者设定固定权重，容易造成解释混乱。较合理的做法是根据课程学习成果明确主要评价对象：若是基础概念掌握，以独立解释为主要证据；若是研究性写作，以论证质量与来源核验为核心；若是人机协作素养，则需要观察学生能否识别工具边界并承担最终决策责任。评分标准应有可见的行为锚点，并通过共同批阅样本提高评分者一致性。

4.4 实施条件与责任分配

重构评价机制需要考虑教师工作负担。增加全过程记录和频繁口头答辩虽可获得更多材料，却未必提高判断质量，反而可能挤占教学时间。较可行的选择是围绕关键学习目标设置有限的核验节点：对基础概念进行独立检验，对高风险结论要求证据核对，对重要推理要求简短解释。这样的设计能够使额外成本与获得的评价信息相匹配。学校应通过教师培训和统一的使用规则，支持工具版本管理、数据安全与评分争议处理。

在学术诚信方面，辨识人工智能生成文本与判断学习成果并不是一回事。Kofinas 等通过英国高校情境中的实验发现，评分者难以稳定地区分人工智能参与程度不同的作品，而真实性较高的任务本身也未能保证能够识别这种参与[19]。由此，评价改革不宜主要依赖文本检测，而应通过任务设计和学习证据降低作品来源不清造成的判断风险。对疑似不当使用情形，应依照明确规则核实，保障学生解释和复核的机会，不能将单一检测结果视为行为认定。

使用条件的公平性也会影响评价结论。不同学生可获得的模型版本、付费功能和设备条件并不一致，允许自由使用工具的成绩可能混入资源差异。UNESCO 关于生成式人工智能教育应用的指导强调人类主体性、隐私和公平等要求[20]。OECD 在 2026 年教育数字化展望中也将学习者能力发展与人工智能辅助表现的区分列为评价的重要议题[21]。对于以能力比较为目的的评价，学校应尽可能统一工具条件；无法使用工具的学生应获得目标等价的替代任务。需要处理学生身份、作业草稿或其他敏感信息时，教师还应限制上传内容并说明数据用途。

5 讨论

5.1 能力替代的解释边界

生成式人工智能对能力形成的影响，取决于工具承担的认知活动与学生保留的目标加工。认知卸载可以降低不必要的操作负担，认知支持也可能促成更有效的反馈循环。但当工具直接完成应由学生练习的核心步骤时，任务质量的提升便可能与能力形成脱节。

该解释也帮助理解现有研究为何呈现不同结果。一方面，实验和元分析提供了人工智能能够改善部分学习结果的证据；另一方面，直接供给答案或缺少学习保护的配置可能产生不利的独立表现。评价结果的差异还可能源于不同研究使用了即时成绩、延迟保持、知识迁移或自我报告等指标。若一项研究测量的是允许工具参与的学习成果，而另一项研究测量的是撤除工具后的独立考试，两者虽然都使用“学习效果”一词，却未必在检验同一能力构念。

5.2 评价重构的教育意义

从评价理论看，本文的核心贡献是把人工智能介入后的表现归因问题纳入学习评价的效度判断。传统课堂也存在同伴协作、教师帮助和家庭辅导等外部支持，因此作品与个人能力并非从来完全一致。生成式人工智能的特殊性在于，它能够快速完成过去通常由学生亲自完成、又被教师用于推断能力的认知活动，从而扩大了“作品可见、过程不可见”的范围。评价改革需要将学生的关键选择、独立解释和迁移表现纳入这一推断链条。

分层证据的选择应依据具体学习目标，重点收集能够解释目标能力的表现。在形成性评价中，过程判断可以帮助教师发现学生何时依赖现成答案；在总结性评价中，必要的独立任务可以支持基础能力认定；在人机协作学习中，核验质量和决策责任则应成为正式评价对象。戴岭等关于教育数字化转型的讨论将“学、教、管、评”视为相互联系的系统[22]，这意味着评价改变应与任务设计、学生工具使用规范及教师支持方式相互衔接。

5.3 局限与研究方向

相关文献的覆盖范围与教学情境影响理论综合的解释范围。现有研究在学段、学科、工具使用方式和结果测量方面存在差异，跨情境比较需要明确这些条件。认知责任的评价还需在具体课程中考察证据采集的可行性、评分一致性及其与独立迁移表现的关系。

未来研究可在具体课程单元中比较直接答案、引导性反馈和不使用人工智能等不同条件下的学习表现，重点考察工具辅助成绩与独立学习能力之间的关系。研究中还应关注学生的先备知识、教师反馈和任务特点，并结合学生解释、作业修订记录及课堂观察，分析不同工具使用方式如何影响学生的认知投入与自主判断。进一步的跨学科、跨学段研究有助于明确能力替代风险的发生条件，并完善相应的学习评价机制。

6 结论

生成式人工智能辅助学习中的关键评价问题，是如何将工具参与下的任务产出与学生真正获得的能力区分开来。能力替代风险产生于工具承担关键目标活动、学生缺少自主判断且评价据辅助作品过度推断独立能力的结合情形。基于认知加工与评价效度理论，本文提出按学习目标组织辅助表现、认知责任与独立迁移三类证据的评价思路。这一思路将评价对象、任务条件与表现证据相互连接，为教师组织教学任务、选择核验节点和解释学习结果提供依据。

参考文献

[1] YAN L, GREIFF S, TEUBER Z, et al. Promises and challenges of generative artificial intelligence for human learning[J]. Nature Human Behaviour, 2024, 8: 1839-1850. DOI: 10.1038/s41562-024-02004-5.

[2] HAN X, PENG H, LIU M. The impact of GenAI on learning outcomes: A systematic review and meta-analysis of experimental studies[J]. Educational Research Review, 2025, 48: 100714. DOI: 10.1016/j.edurev.2025.100714.
[3] KESTIN G, MILLER K, KLALES A, et al. AI tutoring outperforms in-class active learning: An RCT introducing a novel research-based design in an authentic educational setting[J]. Scientific Reports, 2025, 15: 17458. DOI: 10.1038/s41598-025-97652-6.
[4] BASTANI H, BASTANI O, SUNGU A, et al. Generative AI without guardrails can harm learning: Evidence from high school mathematics[J]. Proceedings of the National Academy of Sciences, 2025, 122(26): e2422633122. DOI: 10.1073/pnas.2422633122.
[5] CHEN S, CHEUNG A C K. Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis[J]. Educational Research Review, 2025, 49: 100737. DOI: 10.1016/j.edurev.2025.100737.
[6] SWELLER J. Cognitive load during problem solving: Effects on learning[J]. Cognitive Science, 1988, 12(2): 257-285. DOI: 10.1016/0364-0213(88)90023-7.
[7] RISKO E F, GILBERT S J. Cognitive offloading[J]. Trends in Cognitive Sciences, 2016, 20(9): 676-688. DOI: 10.1016/j.tics.2016.07.002.
[8] CHI M T H, WYLIE R. The ICAP framework: Linking cognitive engagement to active learning outcomes[J]. Educational Psychologist, 2014, 49(4): 219-243. DOI: 10.1080/00461520.2014.965823.
[9] PANADERO E. A review of self-regulated learning: Six models and four directions for research[J]. Frontiers in Psychology, 2017, 8: 422. DOI: 10.3389/fpsyg.2017.00422.
[10] WOOD D, BRUNER J S, ROSS G. The role of tutoring in problem solving[J]. Journal of Child Psychology and Psychiatry, 1976, 17(2): 89-100. DOI: 10.1111/j.1469-7610.1976.tb00381.x.
[11] SHUTE V J. Focus on formative feedback[J]. Review of Educational Research, 2008, 78(1): 153-189. DOI: 10.3102/0034654307313795.
[12] HATTIE J, TIMPERLEY H. The power of feedback[J]. Review of Educational Research, 2007, 77(1): 81-112. DOI: 10.3102/003465430298487.
[13] BARCAUI A. ChatGPT as a cognitive crutch: Evidence from a randomized controlled trial on knowledge retention[J]. Social Sciences & Humanities Open, 2025, 12: 102287. DOI: 10.1016/j.ssaho.2025.102287.
[14] DENG R, JIANG M, YU X, et al. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies[J]. Computers & Education, 2025, 227: 105224. DOI: 10.1016/j.compedu.2024.105224.
[15] KANE M T. Validating the interpretations and uses of test scores[J]. Journal of Educational Measurement, 2013, 50(1): 1-73. DOI: 10.1111/jedm.12000.
[16] 沈书生, 祝智庭. ChatGPT 类产品：内在机制及其对学习评价的影响[J]. 中国远程教育, 2023, 43(4): 8-15. DOI: 10.13541/j.cnki.chinade.20230223.001.
[17] BLACK P, WILIAM D. Assessment and classroom learning[J]. Assessment in Education: Principles, Policy & Practice, 1998, 5(1): 7-74. DOI: 10.1080/0969595980050102.
[18] BARNETT S M, CECI S J. When and where do we apply what we learn? A taxonomy for far transfer[J]. Psychological Bulletin, 2002, 128(4): 612-637. DOI: 10.1037/0033-2909.128.4.612.
[19] KOFINAS A K, TSAY C H H, PIKE D. The impact of generative AI on academic integrity of authentic assessments within a higher education context[J]. British Journal of Educational Technology, 2025, 56(6): 2522-2549. DOI: 10.1111/bjet.13585.
[20] MIAO F, HOLMES W. Guidance for Generative AI in Education and Research[M]. Paris: UNESCO, 2023. DOI: 10.54675/EWZM9535.
[21] OECD. OECD Digital Education Outlook 2026: Exploring Effective Uses of Generative AI in Education[M]. Paris: OECD Publishing, 2026. DOI: 10.1787/062a7394-en.
[22] 戴岭, 胡姣, 祝智庭. ChatGPT 赋能教育数字化转型的新方略[J]. 开放教育研究, 2023, 29(4): 41-48. DOI: 10.13966/j.cnki.kfjyyj.2023.04.005.