

生成式人工智能推动教育评价转型

◆ 蒋慧芳 曾文婕

[摘要] 生成式人工智能推动教育评价转型,呈现五个转向。评价理念从“甄别诊断”转向“素养发展追踪”,评价主体从“教师主导”转向“人机共生型评价共同体”,评价内容从“单一知识”转向“综合素养”,评价方法从“标准化测验”转向“人机协作的多元方法”,评价结果从“静态等级划分”转向“个性化学习发展轨迹与反馈生成”。面对碎片化数据难以反映素养整体、人机协同角色边界模糊、“拼盘式”活动窄化素养内容、评价方法综合应用存在壁垒和教育决策过度依赖静态评价结果等转型过程中的挑战,需要构建素养发展图谱、推动人机协同分工、设计进阶式情境活动、赋能教学评一体化和生成可视化过程叙事。

[关键词] 生成式人工智能;教育评价;评价转型;人机协同

[中图分类号] G40-058.1 **[文献标识码]** A **[文章编号]** 1002-4808(2025)08-0041-08

中共中央、国务院印发的《深化新时代教育评价改革总体方案》指出,“创新评价工具,利用人工智能、大数据等现代信息技术,探索开展学生各年级学习情况全过程纵向评价、德智体美劳全要素横向评价”。《教育强国建设规划纲要(2024—2035年)》要求,建立基于大数据和人工智能支持的教育评价和科学决策制度,深化教育评价改革。2025年4月,《教育部等九部门关于加快推进教育数字化的意见》指出,全面推进智能化,赋能教育评价改革。面对纸笔测试难以评价批判性思维等高阶能力、终结性考核遮蔽学生动态发展轨迹等问题,以ChatGPT, DeepSeek, Sora为代表的生成式人工智能提供破解问题的新的可能。生成式人工智能是依托深度学习算法与复杂神经网络,通过模拟人类认知机制实现创造性内容生产的技术,主要具有多模态内容生成、逻辑推理、情境适应性交互和自动化学习等特征。该技术带来的教育评价智能化变革,不局限于工具理性的技术迭代与操作层面的局部评价技术改良,而是通过对已有评价要素的涵括式超越,推动教育评价系统性转型,进而助力评价改革走出“表层数字化”困境。立足教育强国建设战略需求,本文聚焦生成式人工智能推动教育评价转型

的实践进路,通过分析其认知涌现、动态诊断、情境理解和启发性内容生成的技术赋能逻辑,探讨教育评价转型的向度、挑战和路径。

一、生成式人工智能推动教育评价转型的向度

以教育评价的“五大基本问题”为分析框架,通过探析为何评、谁来评、评什么、如何评和评后怎么用,揭示生成式人工智能推动教育评价转型的向度。

(一) 评价理念转变:从“甄别诊断”到“素养发展追踪”

在20世纪初兴起的教育测量运动推动下,“评价即测量”影响深远。教育评价聚焦“测量学生学习结果”,依赖测验、量化指标进行等级划分和横向比较,追求客观性与可比性。过度强调结果评价,导致三大问题:一是评价维度窄化,将素养简化为可量化的知识与技能指标,忽视高阶思维和情感态度等维度;二是评价功能异化,过度关注分数名次,致使短期的“育分”功能遮蔽长期的“育人”价值;三是个性化支持缺位,教师疲于应付重复考评任务,难以为每位学生提供个性化支持和指导。

蒋慧芳/华南师范大学教育科学学院特聘副研究员(广东广州 510631);曾文婕/华南师范大学教育科学学院教授、博士生导师(广东广州 510631)。

“素养发展追踪”理念强调超越单一甄别诊断,通过持续记录、分析学生在真实情境中多维素养的动态发展过程及其相关证据,促进学生个性化成长。生成式人工智能的智能对话、动态响应和多模态分析等功能,为学生素养发展的个性化追踪提供支撑。通过多轮次、情境化对话交流,采集学生认知、表达和协作等多方面数据。接着,对数据进行初步分析,识别素养发展趋势和潜在证据链,为学生自我反思和评价提供依据,帮助教师洞察素养培育薄弱环节和个体差异。教师对智能分析结果进行验证和解读,并综合考量学生实时素养表现与预设素养发展目标之间的差距,调整教学策略,为学生提供个性化素养进阶方案。

(二)评价主体协同:从“教师主导”到“人机共生型评价共同体”

教育评价长期依赖“教师主导”模式,由教师独立完成评价数据采集、分析和反馈。单一评价主体容易导致教师负担过重、学生主体性被遮蔽和评价反馈滞后等问题。生成式人工智能以智能体的形态代理评价导师、数据分析助手等角色,在与人类主体的持续交互中实现双向发展,由此构建“教师-学生-同伴-智能体”共生型评价共同体,并通过四维互动提升评价效能。一是智能体与学生互动,提供个性化评价支持,赋能学生评价能力增值。首先,利用自然语言处理与深度学习技术分析多模态数据,实时评价学生认知水平与情绪状态,提供伴随式个性化反馈。学生对反馈的采纳情况作为训练数据,持续优化智能体的诊断模型。其次,运用思维链可视化技术显性化评价思维过程,为学生提供评价示范,引导其通过对话、追问与反思发展评价能力。二是智能体与教师互动,优化教师评价实践。智能体辅助教师完成学生表现分析、作业批改和评语生成等工作,同时记录教师对其生成内容的修改痕迹,用于微调模型。教师则将更多精力投入价值引导、评价意义诠释与教育决策优化。三是智能体作为技术中介支持协商式评价。教师与智能体协同制订评价方案,再整合学生建议完善方案。师生还可以共同解析智能体生成的评价标准,就其内涵和实施达成共识。协商确定的标准自动同步至智能体知识库,便于后续评价调用。

四是智能体支持生生互动,提升同伴互评质量。智能体提供结构化评价框架、设定评价指标,实时分析互评数据并检测评价偏差,优化同伴互评过程。

(三)评价内容重构:从“单一知识”到“综合素养”

当前,知识点测试仍占重要地位^[1],评价内容长期聚焦于学生对知识的记忆与复现能力,导致学生陷入解题技巧训练,难以把握学科知识的系统性和知识点之间的关联性,进而使学科间的知识融通及其育人价值被遮蔽。^[2]

生成式人工智能凭借语义理解、多模态分析和情境交互能力,创设贴近现实世界的开放问题情境,在人机交互中分析学生的思考踪迹和问题解决过程,使评价不仅仅考查知识点掌握水平,更聚焦学生批判性思维、问题解决等素养表现。^[3]评价内容向“综合素养”转型,需要经历“素养解构与情境化重组”的过程。素养解构指通过人机协同将素养解构为可观测、可测量的具体维度和行为指标。教师依据已有知识经验建构结构性、框架性和粗粒度的素养指标,接着由智能体生成条目性、颗粒化、细粒度的数据指标,进而通过协商对话形成素养评价指标体系。^[4]素养解构并不意味着对单一素养指标进行割裂式评价,而是解决评价内容维度模糊的问题,便于评价学生在各个素养指标的增值情况。在真实或拟真的任务情境中,学生需要重组解构后的素养指标解决问题,评价内容也从静态的指标维度转变为动态的复杂问题解决过程中所体现的素养。例如,将学生的合作素养解构为愿景认同、责任分担和协商共进三个指标,为评价学生的素养表现提供观测点。^[5]接着,在Sora与VR技术融合构建的动态沉浸式环境中,学生合作完成实验任务。虚拟场景的动态性和情境多样性要求学生灵活重组、综合运用相关素养,不断调整策略、重新分配角色并协商新的解决方案。在此过程中,通过动作捕捉设备、语音记录仪等实时采集学生多模态行为数据,结合Sora可视化呈现学生综合运用素养完成实验任务的全过程。

(四)评价方法革新:从“标准化测验”到“人机协作的多元方法”

标准化测验能快速收集和比较数据、减少主

观偏差,具有较强的客观性。但若仅凭测验判断学生学习水平,容易忽视个体差异性与情境多样性,难以全面评价学生的综合素养。

利用生成式人工智能的动态内容生成、多模态数据融合与个性化反馈能力,开发与应用多元评价方法。由多种方法获得的评价结果相互校准和补足,形成较为全面的评价报告和改进建议。一是人机协作的测验,指生成式人工智能、教师与学生协同优化传统测验。首先,结合师生输入的提示词生成情景创新型测试题,提高题目生成的效率和质量。^[6]其次,根据学生学习情况与能力水平动态调整试题难度和区分度,实现自适应测试,使测验反映学生真实学习水平。最后,采集学生在测验中的答题时间序列、修改痕迹等细粒度数据,精准分析错因,提供个性化指导。二是人机协作的对话式评价,通过“任务参与、智能体分析、反馈提供、反馈接收、与智能体研讨、修订内容”的循环流程,将评价深度嵌入学习过程。^[7]学生既作为“学习者”根据智能体反馈调整表现,又作为“评价者”批判性审视智能体输出内容的合理性。具体而言,当学生与代理“虚拟同伴”的智能体交互时,需先输入问题及初步解决方案,智能体通过提示性反馈引导学生修正方案,而非直接给出答案。学生在接收反馈后进行批判性分析,通过追问、验证与澄清等对话策略判断建议有效性或对内容合理性提出质疑,进而推动新一轮的分析与研讨,直至形成共识并修订内容。三是人机协作的表现性评价,即师生与多智能体系统动态协作,在真实或拟真情境中运用评价标准对学生完成复杂任务中的表现作出判断。多智能体系统含四类智能体。目标制订智能体生成语义精准、内容完整的评价目标,表现性任务设计智能体负责设计真实情境任务并提取可观测的表现特征,评价标准设计智能体构建素养导向的层次化评价标准,评价标准审核智能体对评价标准与素养目标的映射关系进行分析,并通过迭代优化机制实现评价标准的动态优化。^[8]

(五)评价结果改善:从“静态等级划分”到“个性化学习发展轨迹与反馈生成”

以静态等级呈现评价结果便于快速了解学生表现水平,但若将学生复杂认知发展过程简化为

分数、排名等指标,过度强调成果而忽视动态学习过程,易导致评价与改进脱节。生成式人工智能对学生的静态评价结果与实时学习行为数据进行整合分析,生成个性化学习发展轨迹与反馈。

其一,生成个性化学习发展轨迹。依托学习管理系统、智能穿戴设备等采集学生多模态学习数据,经清洗与整合形成全息数据集。接着,利用生成式人工智能语言处理、情感计算和多模态分析能力,对数据进行关联分析,揭示认知、情意、社会互动和自我调节等多维度素养间的关系,帮助教师在立体多维的素养网络中定位学生发展坐标。在此基础上,融合全息数据与关联分析结果,通过时间轴呈现学生在不同学习阶段和领域的连续性成长证据,精准标注关键学习事件中的素养增长点与发展特征,将隐性学习历程转化为可追溯、可解读的个性化学习轨迹。

其二,动态生成个性化反馈。若评价后仅提供静态分数或等级,缺乏深度分析和解释,学生难以理解表现成因并改进。生成式人工智能在提供评价等级的同时,通过深度理解学生与智能体会话周期内的交流信息,提供个性化反馈。一是回溯式反馈。分析学生学习轨迹,解释能力发展遭遇瓶颈的原因。二是前瞻性反馈。将诊断结果转化为下一步改进方案,前瞻未来学习路径。要使反馈真正发挥作用,关键在于学生如何理解并应用反馈改进学习。如果学生主动验证智能反馈的可靠性,或积极引导反馈过程,如输入评分标准让生成式人工智能扮演不同角色进行多角度反馈,学生更有可能根据反馈改进学习。若学生高度依赖外部反馈,不加批判地采纳建议,则难以将反馈转化为自我内部认知与决策。^[9]针对后者,教师应提供提问框架、反馈验证清单等工具,帮助学生利用反馈促进认知调整与能力提升。

二、生成式人工智能推动教育评价转型的挑战

生成式人工智能在推动教育评价理念、主体、内容、方法和结果转型的过程中,面临五大挑战。

(一)碎片化数据难以反映素养整体

学生素养形成与发展是复杂、连续、多维度相互交织的动态过程,生成式人工智能为实现

“素养发展追踪”理念提供可能，但其技术逻辑客观上使评价数据呈现碎片化特征，难以反映素养的整体性。其一，对话记忆有限导致数据记录碎片化，连续的素养发展轨迹被分割为离散片段。生成式人工智能具备较强的上下文理解和多轮对话能力，但其对话窗口有词元限制。当用户输入的提示词超过阈值，生成式人工智能可能无法读取之前的内容或仅读取部分信息，如标题和首段。^[10]即使是上下文窗口达百万级词元的模型，也可能因计算成本过高或无关信息干扰，选择性忽略中间位置的重要信息。^[11]这就导致当追踪长期、复杂的学习过程时，关于同一学生、同一素养发展的连贯记录被拆分成多个离散的对话片段。其二，离散数据割裂素养维度的内在关联。生成式人工智能基于海量数据库生成的内容是各类离散信息的集合。若信息缺乏整合框架，学生认知、情意、社会互动等素养维度之间的交互易被忽视，较难还原素养发展全貌。

（二）人机协同角色边界亟待厘清

人机协同过程中面临权责分配失衡与角色冲突等多重困境。一是智能技术的工具理性偏执导致教育评价主客关系倒置。在评价过程中过度追求效率与量化指标，可能会使评价主体产生技术依赖。智能体从辅助工具转变为隐性决策主体，师生将评价数据收集和分析、评价指标建构、评价决策权力让渡给技术，成为评价结果的被动接受者。^[12]二是师生过度依赖智能评价结果，产生能力退化隐忧。师生对智能评价的盲目信任，忽视评价过程的批判创新思维，导致教师在评价设计、实施和解读等方面的专业能力逐渐退化，^[13]学生的自主思考能力、创新能力降低。^[14]三是人机责任归属不明。当前评价标准不健全、反馈机制缺失及算法偏见影响评价科学性和客观性。若依据机器评价结果进行教育决策，容易引发机器评价的责任归属问题，即评价后果该由机器还是师生承担等。^[15]

（三）“拼盘式”活动窄化素养内容

素养具有情境性与整合性，需要通过学生在真实性活动或表现性任务中的综合表现观测。当前部分评价实践未能充分发挥生成式人工智能优势，形成的“拼盘式”活动缺乏情境性与连续性，难以承载素养评价内容。其一，“拼盘式”

活动常将学生在单一活动或活动某个侧面中表现的“活动素养”，直接等同于学生的整体素养。整体素养是指“作为一个整体人的综合素养或全面发展的素养”^[16]。将局部的、特定情境下的“活动素养”视为整体素养，是对素养评价内容的窄化和简化。其二，将不同活动中观测的“碎片化”素养表现简单累加推断整体水平，割裂素养内容结构。单一活动情境仅能激活部分素养维度，如知识问答侧重记忆再现、小组讨论聚焦协作能力。若将不同活动中的素养简单相加，难以准确评价学生的整体素养。其三，“拼盘式”活动之间缺乏内在连续性和逻辑关联。单一、间断的活动评价仅能捕捉特定时间点的静态表现片段，难以反映素养在真实情境中渐进发展、迁移与深化的过程。

（四）评价方法综合应用存在壁垒

评价方法的革新并非否定或排斥已有方法，而是依托生成式人工智能赋能师生创新性地运用已有方法、创造新方法，并在课堂综合应用，全面深入把握学生学习过程与结果。然而，多元评价方法的综合应用存在困难。其一，评价方法选用较为随意，为了多样而多样。义务教育课程标准提出，应根据不同情况综合选用适切评价方法，如语文课标要求“综合运用多种评价方法，增强评价的科学性、整体性”^[17]。但实践中，教师缺乏对“为何而评”的深入思考，难以结合具体评价目标、学习内容和学生学习特点综合选用适当的方法，也很难在有限的课堂时间和繁重的教学任务中协调多种评价方法。其二，评价方法游离于教学过程。评价被视为独立于教学之外的事件，而非嵌入教学过程的动态环节。所谓“综合运用”，往往只是在学期末等特定时间点堆砌多种方法以满足要求，而非贯穿课前、课中和课后的有机整合，难以发挥评价对学习的动态引导作用。

（五）教育决策过度依赖静态评价结果

评价结果呈现方式虽从“静态等级”转向“个性化学习发展轨迹与反馈”，但教育决策者在实践中仍难以摆脱对静态评价结果的路径依赖，主要原因有三个方面：一是长期形成的评价思维惯性，使分数和等级等量化结果成为判断学习成效的主要依据；二是过程数据解读存在困难，海

量数据虽蕴含丰富的个体发展信息，却因体系庞大、信息过载，且缺乏有效的整合、提炼与叙事化表达，导致决策者较难理解内在逻辑、洞察学生真实的学习难点和成长轨迹，使评价结果的育人增值效应难以充分释放；三是智能素养不足，面对复杂过程数据，部分决策者缺乏必要的数据解读、可视化理解与叙事化思维等智能素养，难以将个性化学习发展轨迹中蕴含的丰富信息，转化为促进学生发展的具体行动。

三、生成式人工智能推动教育评价转型的路径

生成式人工智能为教育评价转型提供新的向度，但在实践中面临诸多挑战。因此，有必要探索系列路径，推动教育评价深度转型。

（一）构建素养发展图谱，贯穿学生学习全过程

素养发展图谱通过结构化、语义化和可视化方式，整合知识单元、学科能力与素养水平，形成多维关联的网状结构，主要功能包括建立知识点层级关系、标注知识点对应的学科核心素养，以及根据学习行为数据实时更新图谱。

其一，在数据维度上，整合不同来源和类型的数据，为图谱提供数据支持。融合生成式人工智能的对话交互数据，智能终端、可穿戴设备等采集的语言、表情、生理信号等多模态学习过程数据，以及在线学习平台、跨学科活动、课堂互动、测试等多场景表现数据。利用生成式人工智能的语言理解和多模态数据处理能力，对复杂数据进行解析，为素养发展图谱构建提供基础。

其二，在结构维度上，建立知识、能力、素养间的结构化映射，形成纵横关联的素养网络。一方面，生成式人工智能通过分析课程标准、教材和教案等资源，提取知识单元并关联相应的素养标签。利用自然语言处理技术对“知识-素养”关系进行语义标注，建立“资源/活动—知识—素养”之间的动态映射关系。^[18]当教师设计“解释当汽车急刹车、转弯时，车内可能发生的现象，讨论系安全带等保护措施的必要性”的活动时，生成式人工智能识别出活动关联的知识点是惯性，并关联到科学态度与责任、物理观念素养。这为后续评价设定了明确目标，即学生能否运用

物理观念解释现象并体现科学责任感。另一方面，通过分析学生的学习行为模式与表现数据，识别并呈现不同素养维度之间的内在关联和协同发展关系。

其三，在时间维度上，图谱贯穿学生学习全过程，伴随学生成长不断丰富。学生学习过程数据实时匹配至图谱节点，既持续丰富图谱信息，又即时评价学生学习水平并提供个性化学习指引。

（二）推动人机协同分工，明确“师-生-机”角色边界

建立“教师-学生-同伴-智能体”评价共同体协同分工机制，明晰各评价主体的功能边界与协作规则，实现技术赋能与教育本真的动态平衡。一是教师作为“专业研判者”，负责专业引导和判断。教师应从重复性评分中解放，关注生成式人工智能难以评价的隐性素养。研究表明，ChatGPT在论证写作评价中虽能精准识别语法错误，但在评价学生逻辑推理和论证深度方面存在困难。^[19]同时，鉴于智能体缺乏对自身“能力边界”的判断，在无法回答问题时依据概率给出模糊答案，可能导致内容同质化、无意义，存在误导风险。教师应切实履行督导职责，对智能评价结果保留最终解释权并及时纠偏。二是学生作为“反思性参与者”，发展评价能力。学生使用生成式人工智能前独立思考，视智能反馈为学习参考而非标准答案，避免过度依赖技术。在获得智能评价后可进一步追问，通过多轮深度交互参与知识建构与意义生成，发展自我反思与评价能力。三是同伴作为“协作支持者”，基于教师的指导和生成式人工智能提供的结构化评价框架，进行同伴互评与反馈。四是智能体作为“辅助执行者”，承担部分评价任务。合理释放智能体的能动性，使其基于预设规则完成评语生成、语法纠错等评价任务，即时提供学习进度提示与个性化反馈。例如，生成式人工智能辅助教师生成评语、为学生的作文提供即时批改反馈。^[20]

在明确评价主体责任分工基础上，进一步构建“人机共评、功能互补”的协同机制。将教师评价、学生自评、同伴互评和智能体评价结合，培养评价主体整合、比较、分析“他人”评价意见的能力，形成更全面的评价信息。^[21]同时，选

择性发挥生成式人工智能的工具性使用、辅助性执行和功能性替代角色,形成“教师指导—同伴互评—智能脚手架”“学生自评—智能伙伴”和“教师督导—智能代理”等交互关系。为确保协同机制有效运行,构建“技术透明、素养提升、伦理约束”三位一体保障体系。一是技术透明化。提高评价数据分析透明度和可解释性,帮助师生知晓评价结果来源与证据。二是智能评价素养培育。教师应具备与生成式人工智能协作设计评价策略、准确解释评价结果、指导学生改进等素养。学生需发展评价判断能力,批判性评价生成式人工智能的内容输出适切性、使用过程合理性,并将其作为反思工具校准自身判断。^[22]三是伦理约束。确保生成式人工智能与人类价值观对齐,师生应掌握基于人类反馈的强化学习方法,利用反馈数据指导模型微调,使其在多个标准上符合社会主义核心价值观,提升其伦理适配性。^[23]

(三) 设计进阶式情境活动,实现评价内容的全景化整合

为克服“拼盘式”活动窄化素养内容等局限,亟须利用生成式人工智能设计并生成进阶式情境活动,横向融合多学科素养维度,纵向追踪素养发展,实现评价内容的全景化整合。在活动过程中,生成式人工智能不仅解析学生的最终解决方案,还能记录人机协同中的对话和问题解决过程,为学生提供全面反馈。

首先,创设沉浸式学习情境。教师输入活动情境描述和具体要求,生成式人工智能据此生成多模态沉浸式虚拟教育场景,作为进阶活动的起点。例如,结合Sora与VR/AR等技术创建高度逼真的虚拟实验环境,让学生在安全的空间中“具身体验”实验过程,实时监控学生的实验状态、操作数据与情绪变化等。^[24]利用ChatGPT创设模拟社交情景,通过虚拟团队项目、在线合作等方式,评价学生沟通、协作和解决冲突的能力。^[25]同时,生成式人工智能分析学生在情境中的交互与表现数据,精准定位其当前素养发展水平。其次,设计进阶式活动。依据各学科学业质量标准的素养进阶要求,人机协同将素养目标细化为具体的活动目标,据此设计难度递增、复杂度提升的连续活动序列,为学生创设富有挑战性与关联性的实践机会和场景,使学生素养在多场景、多

任务中得以整体性观测。当学生在虚拟教育场景中完成连续进阶活动时,其所展现的不仅是单一技能,更是知识迁移、创新思维和团队协作等素养要素在多场景中的整合性表现。最后,生成式人工智能综合分析学生在进阶式活动中产生的连续多源证据,生成涵盖多维度素养表现的分析报告,支持教师对学生素养发展水平进行全景式评价与判断。

(四) 赋能教学评一体化,使评价方法嵌入教学过程

生成式人工智能赋能教学评一体化,使评价嵌入教学各个环节,解决评价方法综合应用难题。其一,注重教学目标、学习活动、评价方法、评价标准和学习标准的一致性,实现教学评一体化。生成式人工智能通过分析课程标准、教学内容和学情数据,辅助教师设定教学目标。基于素养发展图谱,生成式人工智能生成匹配目标的学习活动矩阵,并根据目标复杂性和学生能力层次,智能推荐相应的学习活动与多元评价方法组合,突破单一测验局限。如针对知识性目标,推荐随堂测验等活动,配套选择性反应评价和书面论述式评价方法;针对技能性目标,推荐探究性活动等,配套表现性评价和个别交流式评价方法。^[26]在此基础上,人机协同构建评价标准。以写作教学为例,教师或学生输入“好的作文有哪些核心标准”“请分析该作文获得高分的关键要素,按重要性排序”等提示词,生成式人工智能反馈写作评价标准,并生成自查学习清单、正例与反例等资源。学生在任务完成过程中时时对照评价标准,审视学习水平,实现边做边评、反思改进,从而将教学目标、评价标准转化为内在的学习意图与学习标准。

其二,将评价方法嵌入教学过程。备课阶段,生成式人工智能分析往届数据,预测当前班级学习难点,辅助教师设计分层教学目标、学习活动和评价方法。课中将评价方法融入学习活动,如在探究活动中运用表现性评价,教师通过观察评价学生表现,并指导学生将自身的探究步骤与生成式人工智能生成的思维逻辑链对比,直观识别思维差异。课后,通过协商式评价,引导学生制订个性化的自我调节学习方案,主动觉察、反思和调节学习过程。

（五）生成可视化过程叙事，推荐学习增值路径

利用生成式人工智能的自然语言处理和跨模态转译等功能，深入挖掘和分析数据，将评价结果以可视化的图像或视频呈现，生动展现数据背后的故事，帮助教育决策者把握学生发展过程。

其一，生成可视化过程叙事。可视化过程叙事指利用生成式人工智能动态整合多模态学习数据，生成蕴含时间维度与因果逻辑的成长故事，并以图表、时间线、故事板、漫画和视频等形式呈现学生成长历程。主要包括分析、叙事、可视化和交互四阶段。^[27]分析阶段，处理原始数据，挖掘对叙事有价值的信息。生成式人工智能整合多源数据，运用检索增强生成技术补充背景知识，提升数据质量和可用性。接着，通过人机交互分析数据所反映的学生学习特征等。叙事阶段，生成连贯的叙事文本。如将学生整个学期的学习发展轨迹转化为时间线叙事，标注关键事件，清晰呈现学生学习痛点、难点或盲点，并提供因果解释。同时，按需调整叙事风格，如生成详细数据分析报告、激励性动画提示或进展概览。可视化阶段，将叙事内容转化为视觉形式。生成式人工智能根据叙事内容智能推荐合适的图表等视觉元素。在图表生成过程中，辅助师生梳理素材间的逻辑关联、纠正错误，并通过优化视觉编码和布局设计提升表达的准确性和美观性。交互阶段，师生与生成的可视化内容进行交互，深入探索和理解数据内涵，实现叙事共创。通过与生成式人工智能实时对话，师生可以深入理解叙事内容，并融入课堂观察、个别交流感受等体验性内容，对叙事进行补充与丰富。

其二，基于过程叙事推荐学习增值路径。教师使用适切的统计分析方法计算增值，真实反映学生在认知、情意、社会互动和自我调节等维度的“增长点”，既判断学生学习已增之值，还推荐学习资源和路径帮助学生实现可增之值。一方面，生成式人工智能生成主题情境式表现性任务和学习进度追踪任务等，以任务驱动学生学习持续增值，帮助素养渐进式发展；另一方面，提供评价标准和进步阶梯图等工具，支持师生、生生之间共建学习目标、共享评价证据，协同助推学习增值行动。

[本文系广东省哲学社会科学规划项目“指向师范生教学能力培养的学本评估策略研究”（项目编号：GD23YJY23）、广东省哲学社会科学规划项目“推动中小学育人方式改革的学习为本评估策略研究”（项目编号：GD23XJY53）研究成果]

[参考文献]

- [1] 雷浩, 崔允灏. 核心素养评价的质量标准: 背景、内容与应用[J]. 中国教育学刊, 2020(3):87-92.
- [2] 鲁子箫. 回归生命立场: 教育数字化转型中的知识观重塑[J]. 教育研究, 2025(1):57-69.
- [3] 武法提, 夏志文, 高姝睿. 以生成式人工智能重塑智慧学习环境: 从要素改进到生态重构[J]. 电化教育研究, 2025(1):54-63.
- [4] 周东波, 赵帅, 肖明. 知识引导智能教育评价: 内涵、机制与路径[J]. 教育研究与实验, 2023(4):118-127.
- [5] 徐冠兴, 魏锐, 刘坚, 等. 合作素养: 21世纪核心素养5C模型之五[J]. 华东师范大学学报(教育科学版), 2020(2):83-96.
- [6] 刘明, 郭烁, 吴忠明, 等. 生成式人工智能重塑高等教育形态: 内容、案例与路径[J]. 电化教育研究, 2024(6):57-65.
- [7] WU TING-TING, LEE HSIN-YU, CHEN PEI-HUA, et al. Integrating peer assessment cycle into ChatGPT for STEM education: A randomised controlled trial on knowledge, skills, and attitudes enhancement[J]. Journal of Computer Assisted Learning, 2025, 41(1): 1-19.
- [8] 王永固, 刘泉, 李晓娟, 等. 基于多智能体与混合检索的学科核心素养表现性评价设计研究[J]. 远程教育杂志, 2025(2):31-44.
- [9] TAM CHOI FUNG ANGELA. Interacting with ChatGPT for internal feedback and factors affecting feedback quality[J]. Assessment and Evaluation in Higher Education, 2024, 50(2):219-235.
- [10] 李焕宏, 薛澜. 生成式人工智能应用的使能型风险规制: 以高等教育应用为例[J]. 清华大学教育研究, 2025(1):68-78.
- [11] JAMES A, TROVATI M, BOLTON S. Retrieval-augmented generation to generate knowledge assets and creation of action drivers[J]. Applied Sciences, 2025, 15(11):1-17.
- [12] 鹿星南, 高雪薇. 人工智能赋能教育评价改革: 发展态势、风险检视与消解对策[J]. 中国教育学刊, 2023(2):48-54.
- [13] 宋兆祥, 司林波. 人工智能赋能新时代教育评价改革的逻辑、边界与路径: 构筑智慧教育评价生态的变革之路[J]. 教育科学研究, 2024(11):59-65.

- [14] 吴龙凯, 程浩, 张珊, 等. 智能技术赋能教育评价的时代内涵、伦理困境及对策研究 [J]. 电化教育研究, 2023 (9):19-25.
- [15] 郑永和, 王一岩, 杨淑豪. 人工智能赋能教育评价: 价值、挑战与路径 [J]. 开放教育研究, 2024 (4):4-10.
- [16] 陈佑清. 在与活动的关联中理解素养问题: 一种把握学生素养问题的方法论 [J]. 教育研究, 2019 (6):60-69.
- [17] 中华人民共和国教育部. 义务教育语文课程标准 (2022年版) [M]. 北京: 北京师范大学出版社, 2022:47.
- [18] 王一岩, 吴国政, 郑永和. 生成式人工智能赋能教育信息科学与技术研究: 新机遇、新趋势、新议题 [J]. 现代远程教育研究, 2024 (6):46-54.
- [19] 冷静, 卢弘焕, 代琳. 生成式人工智能赋能批判性思维测评: 基于ChatGPT的应用实验 [J]. 现代远程教育研究, 2024 (6):102-111.
- [20] 夏静芳, 赖伊玲. 生成式人工智能辅助教学的可能与路径: 以中小学语文作文教学为例 [J]. 中小学管理, 2025 (2):36-40.
- [21] 李毅, 郑鹏宇, 张婷. ChatGPT赋能教育评价变革的现实前提、作用机理及实践路径 [J]. 现代远程教育, 2024 (3):9-17.
- [22] BEARMAN M, TAI JOANNA, DAWSON P, et al. Developing evaluative judgement for a time of generative artificial intelligence [J]. Assessment & Evaluation in Higher Education, 2024, 49 (6):893-905.
- [23] 孙立会, 周亮. 生成式人工智能素养: 概念演变、框架构建与提升路径 [J]. 现代远程教育, 2025 (1):11-21.
- [24] 曾明星, 廖柏林, 覃遵跃. 从ChatGPT到Sora: 生成式人工智能如何重塑深度学习场景 [J]. 远程教育杂志, 2024 (6):11-23.
- [25] 严奕峰, 丁杰, 高赢, 等. 生成式人工智能赋能数字时代育人转型 [J]. 开放教育研究, 2024 (2):42-48.
- [26] 查普伊斯 简, 等. 促进学习的课堂评价: 做得对, 用得好 [M]. 赵士果, 译. 2版. 上海: 华东师范大学出版社, 2021:78.
- [27] HE YI, XU KE, CAO SHIXIONG, et al. Leveraging foundation models for crafting narrative visualization: A survey [J]. IEEE Transactions on Visualization and Computer Graphics, 2025:1-20.

Generative Artificial Intelligence Promoting the Transformation of Educational Evaluation

JIANG Huifang, ZENG Wenjie

Abstract: Generative artificial intelligence is promoting the transformation of educational evaluation, including five dimensions: evaluation concept shifts from “sorting and diagnosis” to “competency progression tracking”, evaluation subject changes from “teacher-dominated” to “human-AI symbiotic community”, evaluation content transitions from “fragmented knowledge” to “comprehensive competencies”, evaluation method advances from “standardized tests” to “multidimensional human-AI collaborative approaches”, evaluation outcome shifts from “static categorization” to “personalized learning trajectories with adaptive feedback”. The transformation of educational evaluation faces many challenges: fragmented evaluation data failing to reflect holistic competencies, role boundaries in human-AI collaboration require clarification, fragmented learning activities narrowing down the content of competencies, difficulties in synthesizing diverse evaluation methods, and educational decision-making being overly reliant on static evaluation results. To address these challenges, systematic reform paths are proposed: constructing competency development maps, clarifying the division of labor between human and AI, designing progressive contextualized activities, empowering the integration of teaching, learning and evaluation, and visualizing learning narratives.

Keywords: generative artificial intelligence; educational evaluation; evaluation transformation; human-AI collaboration

(责任编辑 蒲丽芳)